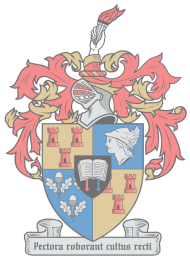
Moral Bioenhancement:
An Ethical and Philosophical Investigation

by
Andrea C Palk



Dissertation presented for the Degree of Doctor of Philosophy in the Faculty of Arts and Social Sciences at Stellenbosch University

Supervisor: Prof Anton A van Niekerk

March 2018

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed, and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety, or in part, submitted it for obtaining any qualification.

Signed – Andrea Palk (March 2018)

Abstract

It has been argued, in the bioethics literature, that traditional forms of moral instruction and development are no longer adequate in equipping humanity to address the urgent nature of the ethical problems facing contemporary societies. Rather than continuing to rely upon these supposedly ineffective methods, we should explore the possibility of biologically enhancing our morality. This would entail both decreasing the likelihood that we would wish to cause harm to others as well as increasing our motivation to do good. The proponents of moral bioenhancement argue that the best way of achieving this would be to isolate which affective dispositions, that are associated with moral traits, are susceptible to biological modification and to biologically enhance these dispositions. However, despite the presence of strong consequentialist arguments in favour of moral bioenhancement, it has elicited a variety of ethical concerns as well as conceptual and practical problems that would have to be addressed for it to become a coherent possibility.

An ethical concern that has been raised in the literature is the concern that moral bioenhancement is wrong, in principle, and regardless of any benefits it could produce, because it risks negatively impacting phenomena that are regarded as intrinsically valuable. In particular, the concern is that moral bioenhancement could impact our moral autonomy, and thus, threaten human morality as such. This concern is based upon the view that the conditions for the exercise of autonomous moral behaviour, and thus morality itself, lie in the deliberation and choice that must be freely made in the face of competing demands. In other words, if it became possible to biologically increase our motivation to do good, thereby increasing the likelihood that we act in a way that is regarded as morally desirable, could our resultant behaviour still be regarded as morally autonomous; or, is morality solely a product of our given, unaltered biological predispositions, working in conjunction with traditional mechanisms of moral education? Will morality as we know it disappear if moral bioenhancement becomes a possibility?

This dissertation contributes towards the literature through a comprehensive review in which particular conceptual, philosophical and empirical problems are addressed, as well as by providing a structured discussion of the practical and theoretical ethical concerns regarding moral bioenhancement. The dissertation includes a substantive definition of moral bioenhancement and makes further independent contributions through the analysis and application of a coherence theory of autonomy to ascertain the status for moral autonomy of various outcomes of moral bioenhancement interventions. From this analysis, a checklist of interventions that could be potentially inimical to autonomy, in terms of their outcomes, is constructed. The conclusion is that in certain cases, moral bioenhancement could produce an increase, rather than a decrease, in the level of autonomy experienced by individuals.

Opsomming

Dit is al meermale in die bio-etiek literatuur beredeneer dat tradisionele vorme van morele onderrig en ontwikkeling nie meer voldoende is om die dringende morele probleme wat teenswoordige samelewings moet aanspreek, die hoof te bied nie. Eerder as om voort te gaan om op hierdie skynbare oneffektiewe metodes peil te trek, moet ons liewer die moontlikheid ondersoek om ons morele sensitiwiteit biologies te verbeter. Dit sal behels dat sowel die waarskynlikheid om kwaad aan ander te doen as die toename van ons motivering om goed te doen, aan die orde moet kom. Die apologete van morele bio-verbetering argumenteer dat die beste manier om laasgenoemde te bereik, sou wees om dié affektiewe disposisies wat geassosieer is met morele kenmerke, te isoleer, te bepaal hoe vatbaar hulle is vir biologiese modifikasie, en om dan hierdie disposisies biologies te verbeter. Ten spyte van sterk konsekwensialistiese argumente ten gunste van morele bio-verbetering, het laasgenoemde moontlikheid ‘n verskeidenheid van etiese vraagstukke sowel as konseptuele en praktiese probleme opgelewer wat aangespreek sal moet word alvorens sodanige verbeteringe ’n koherente moontlikheid kan word.

’n Etiese probleem wat in die literatuur vermeld word, is die vraag of morele bio-verbetering nie miskien as sodanig (in beginsel) verkeerd is nie - ongeag enige voordele wat dit kan oplewer - bloot op grond van die feit dat dit negatief sal impakteer op verskynsels wat inherent waardevol is. Hierdie besorgdheid is veral die vraag of morele bio-verbetering ’n beduidende effek sou kon hê op ons morele outonomie, en dus ’n bedreiging vir menslike moraliteit as sodanig sou kon inhou. Hierdie vraagstelling is gebaseer op die beskouing dat die voorwaardes vir die uitoefening van outonome morele gedrag, en dus van moraliteit as sodanig, geleë is in die deliberasie en keuses wat vryelik gemaak moet kan word ten aansien van kompeterende eise. Met ander woorde: indien dit moontlik is om ons motivering om goed te doen, biologies te verbeter, en om daardeur die waarskynlikheid dat ons op ‘n manier sal optree wat moreel wenslik is, te verhoog, is die vraag of ons resulterende gedrag steeds as moreel outonoom beskou sal kan word. Of, moet ons in so ’n geval, argumenteer dat moraliteit suiwer ’n produk is van ons gegewe, onveranderde biologiese disposisies wat slegs saamwerk met die tradisionele meganismes van ons morele opvoeding? Kortom: sal moraliteit, soos ons dit ken, verdwyn indien morele bio-verbetering ’n moontlikheid word?

In hierdie verhandeling is die gevolgtrekking dat die vlak van bedreiging vir morele outonomie wat morele bio-verbetering inhou, afhang van ’n aantal faktore, wat die aard van die intervensie en die interpretasie van die betekenis van outonomie, insluit. Die argument word verder ontwikkel dat, in sekere gevalle, morele bio-verbetering ’n toename, eerder as ‘n afname, in die vlak van outonomie wat individue ervaar, kan meebring.

Acknowledgements and Dedication

I would firstly like to extend my gratitude towards the National Research Foundation (NRF) for their generous financial assistance towards this research. In addition to the three years of financial assistance that I was awarded, the NRF also provided me with a further travel grant to present my work at a conference in Prague in July 2017 for which I am extremely grateful.

I would also like to extend my heartfelt thanks towards my supervisor and mentor, Professor Anton van Niekerk, who has challenged me, always in a supportive manner, throughout the three years that I have been engaged in writing this dissertation, to find my own voice. It is his confidence in my abilities that has enabled me to have the courage to make my own unique contribution to the literature. In addition, Professor van Niekerk's comprehensive, nuanced and timely feedback on my work, despite the demanding nature of his own responsibilities and work-commitments, is highly appreciated.

On a personal level, thanks must go to Debbi for her support, love and understanding, which gave me the freedom and peace of mind to be able to focus on my writing without any distractions. Finally, I would like to thank my parents, Lawrence and Valerie, for their unwavering support in every aspect of my life. Without their encouragement and affirming presence in my life this doctorate would not have been possible. I therefore dedicate my work to them.

Table of Contents

Chapter 1 – Introduction

1.1 Introduction and overview

The desire to mould the self and improve upon the given is a task with which humanity has occupied itself since the recording of human activities commenced. This drive to improve the human condition has given rise to the creation and development of technological capabilities that enable the innumerable advantages that characterise life in the contemporary milieu. In addition, the development of technology has enabled not only the transformation of the physical world in accordance with human needs, but has also afforded a greater degree of control over human physiology itself. This is reflected in the ever-increasing advancements made in the field of medical science, particularly in the treatment of illness and disease. The successes of medical science have also introduced the possibility of moving beyond the treatment of disease to improving upon the perceived flaws in human physiology. We are now faced with the possibility of extricating human evolution from the contingencies of natural selection and being able to take control of the process ourselves, through biologically improving our cognitive, affective and physical functioning; perhaps even our moral functioning. The possibility of human bioenhancement is, however, fraught with ethical problems and concerns. Whilst there are relatively uniform concerns that may be directed at proposed bioenhancements in all three of these areas of human functioning, possible bioenhancements of affective or moral functioning invoke their own unique set of concerns. It is in this latter area that my research focus is situated; namely, the ethical and philosophical concerns associated with the biological enhancement of morality, which will be referred to from hereon as moral bioenhancement.

The possibility of moral bioenhancement has been suggested as a way of making ourselves morally better; not only in terms of decreasing the likelihood that we would wish to cause harm to others, but also in terms of increasing our desire to do good. In other words, the goal would be to make us care more about, not only other human beings, but also about the world in general. The proponents of moral bioenhancement argue that the best way of achieving the above goals would be to isolate which affective dispositions, that are associated with moral traits, are susceptible to biological modification and to biologically enhance these dispositions.

The idea of moral bioenhancement was first introduced in 2008 in two seminal articles in the *Journal of Applied Philosophy*. These articles established Thomas Douglas and his colleagues from Oxford University, Ingmar Persson and Julian Savulescu, as the dominant proponents of

moral bioenhancement. As is frequently the case with any novel idea or proposal, those who initiate the debate also set its parameters to a certain extent. This has also been the case with the moral bioenhancement problematic in that virtually all subsequent publications have responded directly or indirectly to the arguments of either Douglas or Persson and Savulescu. Much of the discussion in my dissertation will therefore engage with the claims of these thinkers, and in particular, with the arguments of Persson and Savulescu. In this introductory chapter, I will thus commence with a brief description of Persson and Savulescu's argument regarding moral bioenhancement in section 1.2. In section 1.3 I will then provide a motivation for the research area that I have chosen to focus on in my dissertation, which will be followed by an explanation, in section 1.4, of the approach that I will take, as well as what my aims are in conducting this research. In section 1.5 I will outline my problem statement and research questions, after which I will conclude the chapter, in section 1.6, with a brief overview of what will be discussed in each chapter.

1.2 Persson and Savulescu's argument in support of moral bioenhancement

Persson and Savulescu commence their argument by observing that for most of human history, individuals cohabited in groups and social cooperatives that were diminutive in comparison with the vast, diverse and pluralistic societies that now characterise life in the twenty-first century[1]. Furthermore, the technology utilised throughout most of human history was relatively rudimentary, in terms of its potential to negatively impact the environment and other human beings. It is only since the industrial revolution, and in particular, in the previous century, that technology has reached the level of advancement that it is now capable of producing major harm on a global scale, or even worse, existential harm. In this regard, Persson and Savulescu's concern is that we are increasingly faced with the realisation that technological capabilities are such, that individuals and small groups with nefarious intentions are now able to effect great harm upon vast numbers of individuals in a relatively easy manner[2]. This potential for catastrophe lies not only in the use of nuclear weapons of mass destruction, or the misuse of life sciences research for the creation and

[1] The overview provided in this section is an amalgamation of Persson and Savulescu's many publications in which they present their justification for moral bioenhancement (2008; 2010; 2011; 2012; 2013; 2015a; 2015b)

[2] In fact, the claim that it is easier to inflict harm than it is to benefit is an integral part of Persson and Savulescu's argument for moral bioenhancement (2008; 2010; 2011; 2012; 2013; 2015a; 2015b). For example, a lone individual is able to create a bomb that will kill thousands, or drive a vehicle through a crowd resulting in maximum casualties with relative ease. Being in a position to save lives is seemingly more difficult as individuals must already be placed at risk of being harmed in order for one to be in a position to save them. In addition, Persson and Savulescu argue that there are more options available in terms of ways in which one can perpetrate harm than ways in which one can benefit. In this regard, they argue that it is easier to negatively impact a system that is operating efficiently or optimally than it is to enhance such a system.

mobilisation of biological weapons, but also in the failure to forestall the potentially devastating effects of climate change by implementing necessary changes at an individual and collective level.

Persson and Savulescu's general concern is that an increase in the risk of potential catastrophe and existential harm is proportional to advancements made in the field of science and technology. On the one hand, scientific and technological advancement cannot, of course, be halted, as it is a source of untold benefit and progress. On the other hand, due to the existential risk posed by the misuse of technology, coupled with the relative ease of perpetrating widespread harm, Persson and Savulescu posit that the matter cannot simply be ignored. A more specific concern that they voice, and one which lies at the heart of their argument, is that human moral psychology evolved in response to extremely different needs and pressures in comparison to what is now required of it. In other words, for most of human history, human moral psychology has been required to address relatively simple moral issues that were characteristic of life in small societal groupings, rather than the complex, seemingly intractable problems and dilemmas that characterise life in twenty-first century societies. In this regard, Persson and Savulescu attribute many of the societal and global problems facing humanity to individual behaviours and the general inadequacies of human moral psychology.

Furthermore, desirable codes of moral behaviour have traditionally been transmitted through various mechanisms such as parental instruction, education, socialisation with peers, or through the inculcation of a moral code in a specific religious or cultural setting. Persson and Savulescu argue, however, that these traditional methods of moral development are no longer adequate in equipping humanity to address the distinct nature of the problems faced by contemporary societies, nor are they able to motivate the sacrifices required of individuals to bring about general improvements in different areas. Thus, the argument is that both our existing moral psychology, in general, and traditional forms of moral development are simply no longer up to the task required of them. Furthermore, Persson and Savulescu posit that existing political forms of organisation are not equipped to address the consequences produced by the failings of human moral psychology.

One of the most interesting – and I would argue, valuable – parts of Persson and Savulescu's argument is their analysis of the specific aspects of human moral psychology, or *common-sense morality*, that they regard as lacking. Firstly, they point out that human beings are biased in terms of their tendency to pay more cognisance to the concerns of the immediate future. Thus, they are unlikely to make the requisite changes to their behaviour that would suffice in addressing events

that will only occur in the more distant future. An area in which this is particularly concerning, pertains to the behavioural changes that all individuals would be required to make in order to prevent the consequences associated with environmental problems such as climate change and general environmental degradation. Furthermore, this issue is compounded by the fact that given the enormity of human populations, and thus, the perceived dilution of responsibility, there is very little impetus for individuals to adjust their behaviour and display the altruism and sacrifice that would be required to ameliorate the consequences of environmental degradation that we now face.

Secondly, Persson and Savulescu draw attention to the fact that the vast majority of the world's population live in conditions of extreme hardship and dire poverty. Despite the fact that it would require minimal involvement and sacrifice on the part of individuals living in relative affluence, to ameliorate such suffering, help in this regard is mostly absent. Persson and Savulescu give a number of reasons for this general lack of aid-giving to the needy, and, in particular, to the developing world. Firstly, most individuals are predisposed to kin-altruism. This refers to the tendency to experience empathy for, and display altruism towards, only those with whom one is directly connected: one's family, friends, acquaintances or general in-group. In other words, we only truly care, in terms of the extent to which caring motivates us to act, for those who bear some connection or relation to us. This also manifests as a tendency to become overwhelmed by, and desensitised to, the suffering of large groups, particularly those located at a considerable distance from us. This tendency is further exacerbated by the dilution of responsibility: the fact that we feel that our individual efforts in beneficence, particularly regarding the issue of poverty, will be negligible. A further complication is the underlying human tendency known as the *act-omission doctrine*. This refers to the intuition that in weighing our sense of moral responsibility, we place a larger weight on the avoidance of harm than on the provision of benefit. In other words, we feel that it is wrong to harm others but we don't necessarily view it as wrong to not benefit. Persson and Savulescu argue in contradistinction to this view, that in terms of the consequences, failure to benefit, in some cases, ought to be considered as equivalent to direct harm.

It is the combination of all of the above intuitions, dispositions and tendencies, coupled with the growing powers of technology to equip individuals to exact immense harms across vast distances and time, that is so concerning to Persson and Savulescu, and, which motivates their argument for moral bioenhancement. What is needed, Persson and Savulescu posit, is to intensify and accelerate the process of moral improvement by taking matters directly into our own hands. In other words, rather than relying solely on traditional mechanisms of moral enhancement, which they regard as

inadequate to be able to address the urgent nature of the problems we face, we must complement these traditional approaches with other possibilities. The possibility that they regard as offering a promising solution to the above-mentioned shortcomings in human moral psychology would be to make physiological improvements to our morality. This, they argue, will be most effectively achieved through isolating which psychological dispositions, that are associated with moral traits, are able to be biologically modified. Whilst neuroscientific research pertaining to moral bioenhancement is still in its nascent stages, scientific research has already identified two moral dispositions with a biological or genetic basis: altruism and our sense of justice, or fairness. Strengthening these two dispositions, Persson and Savulescu argue, will increase the motivation of individuals to act in a way that they know, in most cases, is good, and will, therefore, address the problems mentioned above in a way that traditional methods of moral development cannot.

1.3 Motivation for the research focus

On a practical level, the possibility of the kinds of moral bioenhancement interventions envisaged in the literature would, of course, be dependent upon the relevant scientific advances being made, and the likelihood of ethical consensus being reached regarding their implementation. However, the exponential rate of progress of scientific and technological capabilities is such that the gap between such aspirations and the development of the means that would enable their fruition is rapidly diminishing. In addition, many of the 'soft' forms of moral bioenhancement suggested in the literature, such as the manipulation of neurotransmitters by means of pharmacological interventions, are already available and are utilised to treat affective disorders. This, therefore, raises the possibility of using these same interventions for enhancement purposes. Due to the fact that moral bioenhancement is a relatively new area of focus, it is vital to clarify and establish an adequate foundation for mediating the ethical issues that are raised by such possibilities, before the more 'hard' forms of moral bioenhancement come to fruition. While it is unlikely that many would agree to undergo moral bioenhancement, for reasons that will be discussed in this dissertation, research in this area could have more applicability for the treatment of mental disorders and behaviour associated with criminal tendencies. This matter will be addressed in the course of my dissertation.

On a theoretical and philosophical level, the moral bioenhancement debate offers many valuable insights. Firstly, the possibility of moral bioenhancement raises important meta-ethical questions regarding the nature of morality itself. Some of the questions elicited by an investigation of moral bioenhancement are: what does it mean to act in a moral manner, rather than simply acting in a

way that has morally relevant consequences; what is the relationship between cognition and emotion in morality; and, is the way in which one becomes and acts morally, something that is, itself, morally relevant? In addition, there are other fascinating questions regarding the relationship between autonomy and morality, such as: to what extent does morality depend upon the possibility of being able to act in an immoral manner, and therefore, to what extent, if any, would moral bioenhancement erode moral autonomy, and thus, morality? This latter question will be the predominant area of investigation in the second half of my dissertation. The moral bioenhancement debate, therefore, enables us to re-examine these age-old areas of philosophical investigation from an entirely different paradigm, and thus, it has reinvigorated important concerns that are deeply indicative of the human condition.

1.4 Research approach and aims

As an area of ethical and philosophical research, there are numerous ways in which one may approach an investigation of the problems associated with moral bioenhancement. It is important to briefly mention some of these possible approaches for various reasons. Firstly, specifying the different ways in which an explication of the problem could be approached will assist in circumscribing the scope of my dissertation. In other words, by pointing out the different ways in which the problem could be approached, or framed, my aim is partly to ensure clarity regarding the approaches, or lines of investigation that I will *not* be taking. Secondly, by pointing out the existence of other ways of approaching the problem, I am wishing to indicate that I am, in fact, very much aware that these different approaches to the problem can be taken, and therefore, that certain important aspects of the problem have not been omitted accidentally, or without consideration. Thirdly, by drawing attention to the existence of alternative ways of approaching the problem, I also wish to indicate that there are other important areas that could be open for further investigation.

In my dissertation, I have aimed to provide an exploratory and comprehensive overview of the problem of moral bioenhancement through a detailed and systematic engagement with the literature. While I have chosen to focus on the ethical and philosophical status of a concern that is prevalent in the literature, namely, the argument that claims moral bioenhancement will impact upon moral autonomy in same way, I have also discussed other practical concerns regarding moral bioenhancement that I take to be relevant to an engagement with the problem. As mentioned above, Persson and Savulescu, have provided a comprehensive argument as to why the possibilities that moral bioenhancement offers, warrant further research. In addition, they have suggested specific

psychological dispositions that they regard as the best potential targets for moral bioenhancement; namely, altruism and a sense of justice. Therefore, one possible approach to an investigation of moral bioenhancement would be to focus extensively on the nature of these dispositions, drawing upon insights from fields such as evolutionary biology and moral psychology in order to ascertain whether or not these dispositions have a genetic basis, as well as their role in morality. Another approach would be to reject the specific dispositions that Persson and Savulescu suggest in favour of other dispositions, or ways of enhancing our morality. One would then be required to provide an argument for why these alternative dispositions would be preferable targets of moral bioenhancement to the ones suggested by Persson and Savulescu. I am not choosing to take either of these approaches, for various reasons.

Firstly, an assessment of the suggested dispositions from an evolutionary biological perspective is an approach that is beyond the scope of this dissertation. Secondly, I am in agreement with Persson and Savulescu that the specific dispositions they have suggested are the most plausible potential candidates for moral bioenhancement. This is not only due to the fact that research indicates that these dispositions may have a genetic basis, and thus, that they may be susceptible to biomedical enhancement, but also, because the nature of these particular dispositions is congruent with a respect for moral pluralism. In other words, if by possessing a sense of justice, what is meant is a respect for the value of fairness in practice, as founded on the notion of the equal moral worth of all human beings, this seems to be a quality that is uncontroversially regarded as good on most accounts of morality. Similarly, altruism or empathy, understood as the ability to regard the welfare of others as morally important, seems to also to enjoy appreciation in most moral codes.

Of course, different accounts of morality would provide different content regarding what would be considered a just or altruistic action. What the nature of this substantive content should be, is, in my opinion, the source of many of the disagreements in the literature. However, this difficulty aside, I believe that the selection of these two dispositions captures something that I take to be an important component of a practical, or non-idealized, morality. In other words, actions that are regarded as just and altruistic are both characterised by the common requirement that in order to be just or altruistic, one must have the ability to take into account the perspective of the other. In a sense, it is impossible to be altruistic or act in a just way if one is unable, or unwilling, to consider an act from the perspective of another. In terms of the specific argument that Persson and Savulescu provide for why we need moral bioenhancement in the first place, I believe that the enhancement of our ability to assume and act upon the perspective of the other is the best starting point for the

discussion. I also believe that this ability is something that comes close to enjoying universal appreciation in most accounts of morality.

Whether the modification of these dispositions, in the way that the proponents envisage, will ever be scientifically feasible, is, of course, a different matter. Furthermore, even if the enhancement of these dispositions becomes scientifically possible, it is not necessarily the case that this would constitute an improvement of our morality. In other words, while a strong sense of justice and altruism may be relevant, or sufficient to compel moral action, it is not necessarily the case that their presence *alone* is sufficient to compel moral action. Morality is a complex phenomenon that requires interaction between various internal components, such as affective and cognitive mechanisms, as well as the ability to interpret and apply relevant contextual information. This latter issue is important for the concern for moral autonomy and will therefore be addressed in my dissertation. While I will briefly address the way in which the issue regarding the scientific feasibility of moral bioenhancement is discussed in the literature in chapter 2, I will make certain idealising assumptions regarding this matter. For the purposes of my investigation, which is ethical and philosophical in nature, I will assume that safe moral bioenhancement *could* be a scientific possibility in the future, given that certain biomedical, and specifically neuroscientific advancements, are made. I am aware, however, that an investigation of the scientific feasibility of moral bioenhancement is a second possible approach that could be taken as a focus for an investigation of moral bioenhancement, and, that the assumption I am making regarding its possibility is a substantial one. I have not chosen to focus on this matter in depth as it does not lie within the ambit of my expertise but is rather an area requiring extensive neuroscientific knowledge.

Regarding the main focus of my dissertation, the concern for the impact that moral bioenhancement may have on moral autonomy, there are also different ways in which this matter may be approached. Once again, one may approach this concern from a neuroscientific paradigm and deny that this is, in fact, a problem, due to evidence that contests the extent to which we possess the kind of autonomy portrayed in the literature, to begin with. Given that this line of argumentation would take my dissertation in an entirely different direction, and not require engagement with my chosen topic, I will not pursue such an investigation. Secondly, one could engage with the problem in an immanent manner by investigating it in detail, and on its own terms, in order to ascertain the legitimacy of the concerns that have been voiced by opponents. In other words, one may assess the concern by using a conception of autonomy that would be regarded as legitimate by the

opponents of moral bioenhancement, and analyse their arguments from within, or, in accordance with this conception. A third possible approach could be to reject the concern outright, on the grounds that these arguments overemphasise the importance of moral autonomy, and autonomy in general.

This third approach leaves little room for engagement with the ethical status of moral bioenhancement itself, as it distils down to a dispute between values. Furthermore, in the case of moral bioenhancement, such an approach is characteristic of a clash between consequentialist arguments that support moral bioenhancement on the grounds of some worthwhile benefit that it will produce, and arguments that are non-consequentialist in nature, and regard our moral autonomy as absolute in value, regardless of any benefits associated with any impacts upon it. Other than noting that this approach can be taken, and briefly discussing it in chapter 4, I will not attempt to provide a solution, as it is an ongoing meta-ethical dispute that I take to be irresolvable. Rather, I will take the second approach – mentioned above – in my investigation and engage directly with the concerns voiced in the literature regarding the potential impact that moral bioenhancement may have on moral autonomy. This approach will require that I investigate and discuss the literature in a detailed manner so as to then synthesize the various arguments and claims, in order to be able to adequately engage with them and offer an independent interpretation and assessment of the problem.

1.5 Research questions and problem statement

The research problem that I will focus on in my dissertation is the concern for human moral autonomy that is posed by moral bioenhancement. In other words, if it became possible to increase our motivation to do good, thereby increasing the likelihood that we act in a way that is regarded as morally desirable, could our resultant behaviour still be regarded as moral; or, is morality solely a product of our given, unaltered biological predispositions, working in conjunction with traditional mechanisms of moral education?

In my dissertation, I will firstly investigate this concern in terms of how it has been elucidated in the literature. Here, the concern is framed and approached in terms of the extent to which morality is dependent upon the possibility of doing wrong. In other words, the prevalent intuition in the literature is that the conditions for the exercise of moral behaviour lie in deliberation and choices that must be freely made in the face of competing demands. Therefore, the concern is that if we have been biologically influenced to make better choices, may these subsequent choices and the

behaviours that they support, still be regarded as moral? Would morality as we know it disappear if moral bioenhancement became a possibility?

I will then utilise the insights of influential theories of personal autonomy to approach the research problem in a different manner. Here, I will conduct an independent analysis of the possible outcomes of moral bioenhancement interventions in terms of their impact upon autonomy. This will require me to elucidate the nature of autonomy, itself, in order to coherently assess whether moral bioenhancement interventions would, in fact, impact upon the former. The research questions that I will, therefore, attempt to answer are: 1) what kind of autonomy is at stake in the moral bioenhancement debate; 2) if moral bioenhancement were to impact autonomy, how would it do so; and, 3) are there any differences between various types of interventions in terms of their possible impact upon autonomy?

1.6 Overview of chapters

In chapters 2, 3 and 4 I will focus on the way in which moral bioenhancement has been discussed in the literature. In chapter 2, I will investigate three interconnected areas that I have identified from a survey of the literature as relevant for my research focus. These areas pertain to how moral bioenhancement is – and should be – defined; what the target/s of moral bioenhancement should be; and, whether or not moral bioenhancement could be a scientific possibility. In chapters 3 and 4, I will then discuss the different arguments that have been made in the literature in opposition to moral bioenhancement. In chapter 3 I will focus specifically on arguments that posit that moral bioenhancement is morally problematic on a practical level, due to potentially negative consequences it will produce at both individual and societal levels. Some of the problems that I will investigate here, range from potential safety risks and implementation and administrative concerns, to concerns for distributive justice and egalitarianism.

I have divided chapter 4 into two parts that I take to be related to each another. These sections address concerns that moral bioenhancement is wrong, in principle, and regardless of any benefits it could produce, because it risks negatively impacting phenomena that are regarded as intrinsically valuable. Here, I will investigate the concern for the way in which both personal qualitative identity and moral autonomy could be affected by moral bioenhancement, and why this would be considered a negative outcome. I have also divided chapter 5 into two parts. In the first part I will present the theoretical component that will inform my analysis in the second part. Theories that will be discussed include hierarchical accounts of autonomy as well as Ekstrom's coherence theory

of autonomy. My aim in this regard is to provide an account of autonomy that is sufficiently rigorous to adequately address the concern for moral autonomy that is posed by moral bioenhancement. I will conclude my dissertation in chapter 6 by synthesising the insights in the afore-mentioned chapters, and by suggesting the areas that warrant further research and ethical discussion. Regarding the level of threat posed to moral autonomy by moral bioenhancement, my conclusion will be that this will depend upon a number of factors, including the nature of the intervention and the interpretation of autonomy. Furthermore, I will argue that in certain cases, moral bioenhancement could produce an increase, rather than a decrease, in the level of autonomy experienced by individuals.

Chapter 2 –Definitions, moral content and scientific feasibility

2.1 Introduction and overview of chapter

In this chapter, I will explore several interconnected and overlapping themes – relevant to the focus of this dissertation – that I have identified in the literature on moral bioenhancement. The discussion is based on a review of approximately 120 publications that have directly addressed the moral bioenhancement problematic since 2008 when the area was first identified in two seminal articles in the *Journal of Applied Philosophy*. These articles established Thomas Douglas and his colleagues from Oxford University, Ingmar Persson and Julian Savulescu, as the dominant proponents of moral bioenhancement. Virtually all subsequent publications addressing the moral bioenhancement problematic have responded directly or indirectly to the arguments of either Douglas or Persson and Savulescu. The focus of this chapter is primarily on conceptual issues related to the identified themes, and the way in which these matters have been addressed by various thinkers in the literature. The various arguments against moral bioenhancement, and some of the responses to these arguments, will be presented in chapters 3 and 4.

In section 2.2 I will investigate one of the most prevalently addressed areas in the literature, namely, definitional issues regarding the term moral bioenhancement itself. This area may be understood as *The problem of competing definitions* as it involves disagreements regarding a topic of fundamental importance, namely, how moral bioenhancement should be defined, and, whether agreement regarding a functional definition is even possible. In section 2.2.1 I will present definitions that characterise moral bioenhancement in terms of its outcomes, as well as more neutral definitions that attempt to avoid notions of improvement or betterment. In section 2.2.2 I will then examine definitions that presuppose that moral bioenhancement will result in moral improvements. The definitions discussed in this section are also more normative in character, due to their specification of the target/s of moral bioenhancement. In this section, I will also discuss the relevance of the distinction between traditional moral enhancement, through non-biological means, and moral biological enhancement, for the ethical status of moral bioenhancement.

In section 2.2.3 I will discuss definitions that utilise notions of species-typical or 'normal' levels of moral functioning. This approach distinguishes between treatment (of pathological levels of moral psychological functioning) and enhancement (above species-typical levels of moral psychological functioning). Finally, in section 2.2.4 I will briefly mention other ways in which definitions may be categorised. These definitions distinguish between broad versus specific and

passive versus active interventions; both distinctions loosely correlate with the distinction between traditional moral enhancement and biological moral enhancement.

In section 2.3 I will discuss the second area of focus which I have termed *The problem of moral content*. This area is the primary source of the problem of competing definitions. The fundamental area of dispute pertains to the question of what makes a person moral. Does morality lie in the actions or behaviour of individuals? In other words, is morality something discernible that can be observed, the kind of person someone is? This seems to be important as morality should be linked to real world outcomes to be meaningful. However, it cannot be all there is to morality as someone may act in a seemingly moral manner due to external manipulation. We can ask then if morality is rather characterised by an internal state. Does morality entail particular attitudes, capacities, intentions or feeling the 'correct' emotions or morally salient motives that compel us to act in a particular way; or, does it lie in the process of cognition or reasoning that we employ to reach particular moral conclusions? Or, is morality informed by a complex conglomeration of all of the above-mentioned factors? It seems likely that this latter possibility would be closest to the truth. This area is not only of theoretical interest, but also, more importantly, it has major implications for practical issues regarding moral enhancement. In other words, these discussions are aiming at answering the question of what we should enhance.

In section 2.3.1 I will discuss the view of the proponents: Persson and Savulescu, as well as Douglas, who have argued that the appropriate target of moral bioenhancement should be moral motivation. In other words, they argue that moral bioenhancement should make us more likely to act in a morally desirable way by strengthening our reasons or feelings for doing so. Due to the emphasis of affective components on our motivation by the supporters, there has been a tendency to frame the debate as one between the role of emotions versus reasoning in morality[3]. This matter, which may be traced back to disputes, in this regard, between Hume and Kant, will therefore be addressed in section 2.3.2. In section 2.3.3 I will then discuss the 'problem' of moral pluralism and the implications that this has for identifying a suitable target for moral bioenhancement. In section 2.3.4 I will briefly discuss John Harris' response to the problem of moral content while in section 2.3.5 I will present interpretations offered by some thinkers who argue that morality is more complex than the emotion/reason dichotomy would imply. I will conclude this discussion of the second problem in section 2.3.6, by briefly discussing the possibility of utilising insights from virtue ethics in order to address the problem of moral content. This approach has been suggested

[3] I will refer to this from hereon as the emotion/reason dichotomy

by several thinkers in the literature as a possible way of overcoming the seeming irreconcilabilities of the above-mentioned approaches.

As mentioned above, the problem of moral content has practical relevance because for moral bioenhancement to be scientifically feasible we would have to know what it is we should attempt to enhance. Furthermore, it would have to be established that the identified target/s are, in fact, susceptible to biological enhancement. The third area of focus that I will therefore address in section 2.4 is *The problem of the science of moral bioenhancement*. The proponents of moral bioenhancement are, not surprisingly, more optimistic regarding the scientific possibility of moral bioenhancement. In section 2.4.1 I will therefore discuss the arguments presented by those who regard moral bioenhancement as a scientifically feasible prospect in the future. This will also include a discussion of the potential targets of moral bioenhancement that have been suggested, as well as some of the interventions that have been proposed. The arguments here draw upon insights, studies and research from a variety of fields in support of their claims.

In section 2.4.2 I will investigate the responses to these claims. Certain thinkers have argued that while there may be a partly biological basis to our morality, moral bioenhancement will not – or *may* not – be a possibility due to the complexity of our moral psychology. In other words, it either won't work, or safety concerns will make it too risky. Some argue that the role that physiological – and more specifically, genetic – factors play in influencing our morality has been exaggerated and that the environment exerts a far stronger influence in this regard. According to this view, it is likely that moral bioenhancement is a scientific impossibility. I will address this matter further in section 3.2 of chapter 3 as it is relevant for the concern that moral bioenhancement is wrong in practice due to potential harms and risks that it poses to individuals.

The fourth relevant area of focus in the moral bioenhancement literature addresses *Why we should or shouldn't morally bioenhance ourselves*. This area includes a variety of arguments that have been given by proponents in support of moral bioenhancement and the responses to these claims lodged by opponents. I will discuss these arguments in chapters 3 and 4.

2.2 The problem of competing definitions

A number of thinkers have explicitly drawn attention to the lack of consensus regarding a definition of moral bioenhancement (Douglas, 2008; Persson & Savulescu, 2008; Beck, 2015; Crutchfield, 2016; Focquaert & Schermer, 2015a; Hauskeller, 2015; Pacholczyk, 2011; Raus et al, 2014; Shook,

2012; Walker, 2009; Jotterand, 2011; Bruni, 2011; Schaefer, 2011; Baertschi, 2014; Lechner, 2014; Chan & Harris, 2011, Verkiel, 2017). As mentioned above, this may be directly attributed to the fact that definitions regarding what would constitute an enhancement of morality are directly informed by deeper meta-ethical disputes regarding the content and nature of morality itself. As argued by Raus et al., these deeper meta-ethical disagreements have resulted in most definitions of moral bioenhancement being "significantly less descriptive and more normative than they are regularly portrayed to be" (2014:263). In other words, definitions are generally not content-neutral but reflect a particular interpretation of what morality *should* consist of.

Such meta-ethical disputes are a long-standing phenomenon with no obvious or imminent solution. Therefore, as pointed out by Raus et al, it is not likely that there will be – or should be, for that matter – a "single and universally agreed upon definition" (2014:272) for moral bioenhancement in the foreseeable future. In light of this, they suggest that it would be completely acceptable for the term, moral bioenhancement, to be used as an *umbrella term,* under which a variety of possible definitions could be grouped (ibid.). Agar also argues that "we can endorse a conceptual pluralism that acknowledges the need for more than one concept of human enhancement to address the hugely varied ways in which technology may alter humans" (2014:369). An important requirement of this approach would be to ensure that one specifies the underlying normative stance that informs the definition one is offering, rather than presenting a definition that contains implicit normative underpinnings as neutral. This will also assist with clarity and ensure that interlocutors are able to avoid talking past one another and engage coherently.

Before I discuss the various definitions that have been offered, it is necessary to provide a definition of biological enhancement in general. Buchanan's definition, which is, in all likelihood, the most extensively cited definition in the enhancement literature, refers to biological enhancement as "a deliberate intervention, applying biomedical science, which aims to improve an existing capacity that most or all normal human beings typically have, or to create a new capacity, by acting directly on the body or brain" (2011:23). The Nuffield Council also provides a good working definition of general enhancement, defining it as "the directed use of biotechnical power to alter, by direct intervention, not disease processes but the 'normal' workings of the human body and psyche, to augment or improve their native capacities and performances, and in that sense, is taken to be beyond therapy" (2013:164). Beauchamp's further specification of the term enhancement is also useful. He posits that in the context of the moral bioenhancement debate, "by the word 'enhancement' the authors seem to mean an intervention – a human initiated improvement in traits,

dispositions, characteristics, or capacities intended to take persons beyond normal levels of human functioning" (2015:346). These three definitions are seemingly similar but can be distinguished somewhat due to their different emphases. All three definitions explicitly use the interpretation of enhancement as improvement. However, Beauchamp and the Nuffield Council's definitions explicitly refer to improvement *above normal functioning*, a contested approach due to the difficulty in identifying what would constitute normal functioning. This approach is nevertheless utilised by a number of thinkers and will be discussed further below.

2.2.1 Outcomes-based definitions versus neutral definitions

In terms of defining moral bioenhancement, however, it is not only the above-mentioned lack of consensus regarding definitions that is striking, but also the dearth of thinkers who actually offer definitions to begin with, that is noticeable when reviewing the literature. Whilst there are an abundance of publications addressing the moral bioenhancement problematic, there are only a handful of substantive definitions offered. The broadest definitions define moral bioenhancement in terms of the outcomes it will produce. In a collaborative article, Savulescu, Douglas and Persson offer this type of definition of moral bioenhancement, defining it as "an intervention that makes it more likely that you will act morally, in some future period, than would have been the case if it were not used. One acts morally when one does the right thing, and for the right reason(s)" (2014:95). Savulescu, Douglas and Persson do, however, draw attention to the fact that providing content regarding what it means to act more morally will depend on one's account of morality and will thus be open to dispute. Pacholczyk offers a similar definition of moral enhancement as involving "a change in some aspect of morality that results in a morally better person" (2011:252). She argues that becoming morally better could, of course, involve a variety of possibilities such as "making people more likely to act on their moral beliefs, improving their reflective and reasoning abilities as applied to moral issues, increasing their ability to be compassionate and so on" (Pacholczyk, 2011:252).

Pacholczyk points out that describing moral bioenhancement in terms of using notions of *moral betterment*, or improvement, for that matter, is, of course, a normative claim. It would remain to be seen whether an intervention does, in fact, result in moral betterment; and, whether this is the case would depend upon one's conception of what moral betterment consists in (Pacholczyk, 2011:253). Simkulet offers a definition which implies a similar view. He defines moral enhancement as "any enhancement that improves the likelihood that a moral agent will achieve his

or her moral goals, where moral goals can be either praiseworthy or blameworthy depending on their intentional content and epistemic character" (Simkulet, 2012:18).

Due to the recognition of this problem, some thinkers attempt to provide definitions or use terminology that explicitly excludes the idea of improvement. Baertschi uses the term neuromodulation (2014), whereas Murphy refers to biomoral modification rather than moral bioenhancement, thus avoiding the issue "of whether the modification actually amounts to an enhancement" (2015:370). Zarpentine attempts to provide a similarly neutral definition, describing moral bioenhancement as involving "direct pharmacological or surgical manipulation of the brain or selection of genetic material conducive to the aims of moral enhancement" (2013:142). Sparrow defines moral bioenhancement as a "deliberate modification…[via 'drug therapies, neural implants, or (perhaps) genetic engineering] in order to make [people] 'more moral'" (2014b:20). While his definition contains an implicit idea of improvement, Sparrow places the phrase 'more moral' in scare quotes in order to emphasise his position that moral bioenhancement, as envisaged by its proponents, would not, in fact, make us more moral. By using terms such as modification, alteration, manipulation or change, the above definitions are indicative of obvious attempts to steer clear of the normative implications associated with notions of improvement. Such terms may also signify that a thinker contests the view that moral bioenhancement would, in fact, be an improvement or enhancement, as is the case with Sparrow.

2.2.2 Target-based, normative definitions

A more common approach in the moral bioenhancement literature is to sidestep the issue of whether or not interventions would result in definitive improvements in morality, and, to simply assume that they would. Any intervention named a moral bioenhancement under this interpretation would therefore, be an improvement, by definition. This approach is characterised by a focus on the target of enhancement. In other words, it defines moral bioenhancement by what it would aim to enhance, or work on. In this regard, such definitions are normative as they provide specific, substantive content in terms of what *should* be enhanced, and thus, they are offering a particular perspective on what constitutes morality or the nature of morality in general. In terms of this approach, De Melo-Martin and Salles have argued that despite the differences in what they would purport to be the targets of moral bioenhancement, these approaches all have in common the fact that moral bioenhancement is seen "as one more means in a continuum of practices whose goal is *moral improvement*, that is, as means of creating *morally better* people" (2015:3).

There are a variety of target-based definitions that have been offered, including both broad and more specific definitions. In terms of wider definitions, Jebari defines moral bioenhancement as altering "a person's dispositions, emotions or behaviour in order to make that person more moral" (2014:253); whereas Crutchfield defines it as "the enhancement of a person's moral attitudes, motivations, or behaviour through biological means" (2016:389). Crutchfield admits that his definition is problematic in that it isn't clear what enhancements in these areas would entail (ibid.). Furthermore, he points out that while his definition could be amended to include "dispositions and emotions…[it is] general enough to capture the notion of moral bioenhancement and specific enough to be useful" (ibid.). Shook also alludes to the variety of possible contenders for enhancement, pointing out that what could be considered "'moral' enhancement ranges from feeling empathic concern to increasing personal responsibility all the way to heightening respect for global fairness" (2012:3). He posits that "moral intuitions, virtues, and rules are not identical around the world" (ibid.) alluding to the fact that what may be regarded as morally positive in one location could be regarded in a less positive light elsewhere. However, he argues that despite the presence of plural conceptions of morality and the complexity of identifying the content/s of morality, there are areas of overlap that can be used to devise a list of potential targets for moral bioenhancement (Shook, 2012:4). Shook posits that moral bioenhancement can be defined as improving functioning in five areas: *moral appreciation, moral decision-making, moral judgements, moral intentions* and *moral will power* (2012:5-6)

There are also thinkers who provide definitions that attempt to narrow down the target/s of moral bioenhancement. Christen and Narvaez define moral bioenhancement as "the endeavour to improve moral *behaviour* in a neuroscientifically informed way" (own emphasis, 2012:25). As mentioned above, Sparrow has argued that in the literature, moral bioenhancement has been described as aiming at modifying "individuals' *behaviour* and *dispositions* in order to make them 'more moral'" (own emphasis, 2014b:20). However, elsewhere he contests whether such interventions would in fact achieve their purported aims; arguing that whilst it is feasible that certain interventions could change:

> behaviour and emotions in ways that we may be inclined to morally evaluate positively, describing this as a moral enhancement presupposes a particular contested, account of what it is to act morally, and also implies that entirely familiar drugs such as alcohol, ecstasy, and marijuana [which alter behaviour and emotions in a similar manner] are also capable of making people 'more moral' (Sparrow, 2014a:24).

Barilan discusses the traditional understanding of enhancement, arguing that if "enhancement means betterment…[then] in this sense, moral enhancement may count as an increase in kind,

charitable, and just *judgements* and *actions*" (own emphasis, 2015:75). In terms of this inclusion of moral judgement and the actions it leads to, this definition identifies three important, and much-discussed components of morality; namely, dispositions or virtues (the terms *kind, charitable* and *just*), reasoning (the term *judgements*) and behaviour (the term *actions*). A similarly astute definition provided by Focquaert and Schermer purports moral enhancement to refer "to interventions that aim to improve moral decision-making and behaviour" (2015a:141). This definition supports their argument that the means one takes to ensure moral behaviour, are morally relevant. In other words, the inclusion of the term *moral decision-making* implies that moral behaviour has a cognitive component, or, at least, that to be considered moral, it *should* be the product of some process of reasoning that leads to morally improved behaviour. However, they admit that this interpretation of moral bioenhancement requires an accompanying "account of how moral decision-making and moral behaviour work" (Focquaert & Schermer, 201a5:141). If one delves further below the surface of Focquaert and Schermer's definition, it becomes apparent that their definition alludes to the schism in views regarding the content of morality as informed either by Humean emotions or Kantian reasoning – the emotion/reason dichotomy – as well as the concern that for moral bioenhancement to be meaningful it would have to result in some discernible difference in behaviour. These points will be discussed in the next section as well as in subsequent chapters.

Moving to even narrower definitions, Douglas emphasises motives, defining moral bioenhancement as resulting in an individual having "morally better *motives*…[where motives refer to] psychological – mental or neural – states or processes that will, given the absence of opposing motives, cause a person to act" (own emphasis, 2008:229). Later, Douglas defines moral bioenhancement as "interventions that will expectably leave an individual with more moral motives or behaviour than [he or she] would otherwise have had" (Douglas, 2013:162). His addition of *behaviour* implies the relatively uncontroversial assumption that to be considered as a moral bioenhancement, an intervention must produce sufficient motivational force to result in an actual change in behaviour. In addition, the inclusion of behaviour alludes to the possibility of moral bioenhancement that could result in improvements to behaviour without any change in motives. Of course, behavioural improvement in the absence of accompanying changes in moral motivation or reasoning would generally be viewed as problematic as it could be regarded as a possible form of behavioural control[4]. Douglas also defines moral enhancement as aiming at "the attenuation of

[4] This is a major concern of certain opponents of moral bioenhancement (for example: Morioka, 2014; Bublitz, 2016; Simkulet, 2012; Jotterand, 2011; Harris, 2014; Sparrow, 2014b). Harris, for example, argues that, contrary to Douglas and Persson and Savulescu's interpretation, enhancements should not be "define[d]…in terms of the intention or the

counter-moral emotions: emotions that interfere with moral reasoning, sympathy and all other plausible candidates for 'morally good motives'" (2013:161). Examples of counter-moral emotions that are extensively discussed by Douglas include "racial aversion and impulses to violent aggression" (2013:161).

Persson and Savulescu define the concept in terms of enhancing primary *moral dispositions* that are argued to have a biological or genetic base. As discussed in the previous chapter, the two dispositions that they have identified as having the greatest impact upon the problems facing 21st century humanity as well as the threat of ultimate harm, are altruism and our sense of justice (Persson & Savulescu: 2012:107). Thus, for Persson and Savulescu, moral bioenhancement, broadly defined, would involve the attenuation of any dispositions that could impinge upon "the central moral dispositions of altruism and a sense of justice" (2015a:348), or, the boosting of these dispositions themselves.

While Douglas focuses on motives and attenuation of counter-moral emotions, and Persson and Savulescu focus on morally relevant psychological dispositions, David DeGrazia, another proponent of moral bioenhancement, focuses on *capacities*. DeGrazia sets out to define moral bioenhancement without reference to concepts of normal functioning. He defines general enhancement as "any deliberate intervention that aims to improve an existing capacity, select for a desired capacity, or create a new capacity in a human being" (DeGrazia, 2014:361) and moral enhancement, more specifically, as "interventions that are intended to improve our moral capacities such as our capacities for sympathy or fairness" (ibid.). Both definitions are worded in such a way that they could include both non-biological and biological moral enhancements. This is a very important point that warrants a brief diversion as it alludes to the distinction between traditional moral enhancement, through non-biological means, and moral biological enhancement.

Traditional moral enhancement refers, of course, to the inculcation of desirable codes of moral behaviour through various mechanisms such as parental instruction, education, socialisation with peers, religious instruction or the internalisation of particular cultural tenets. Those who support moral bioenhancement, such as DeGrazia and Persson and Savulescu, often argue that there is essentially no morally relevant difference between traditional moral enhancement and moral

motivation of those who produce them but rather in terms of their effect" (2014:372). Defining moral bioenhancement in terms of effects or outcomes will ensure that mere behaviour modification cannot be considered to be an enhancement. This issue is related to the concern for the impact of moral bioenhancement on moral autonomy and will therefore be discussed further in chapters 4 and 5

bioenhancement; they both aim at the same end result and have an equally enduring influence on the individual in question (DeGrazia, 2014; Persson and Savulescu, 2012). Thus, by providing a definition that collapses the difference between the two forms of moral enhancement and presents moral enhancement as occurring on a continuum of sorts with traditional enhancement at one end and moral bioenhancement at the other end, DeGrazia is clearly attempting to provide legitimacy to the project of moral bioenhancement. In other words, the argument is that if we accept traditional moral enhancement then we ought to accept moral bioenhancement.

In fact, the project of moral bioenhancement is deeply informed by the implications of this distinction. The reason why moral bioenhancement is presented as something that requires urgent consideration is due to the underlying belief that traditional moral enhancement is simply not effective enough, on its own, to address what is required of it (Persson & Savulescu, 2012). The argument is that we need something more effective to complement traditional moral enhancement. Those who argue against moral bioenhancement respond in a variety of ways, ranging from claims that traditional moral enhancement is sufficient, to arguments that the means we take to improve ourselves morally, matter. In other words, moral worth lies not only in the end-goal of morality, namely, moral actions, but also in how one arrives at this end-goal. According to this view, a moral outcome that is the product of traditional moral enhancement or education by way of deliberation is of more moral worth than one that is the product of a biological intervention. I will return to this important matter in subsequent sections and chapters.

2.2.3 Moral bioenhancement versus moral treatment

Returning to DeGrazia's definition of moral bioenhancement, he provides further content to his understanding of moral enhancement, arguing that it would encompass improvements in motives and insight which would lead to improved behaviour (2014:263). Furthermore, he provides an extensive list of moral defects that impact upon motives and insight, arguing, in a similar manner to Douglas, that the attenuation of such defects would also rightly be considered a moral bioenhancement (DeGrazia, 2014:364). While DeGrazia defines moral enhancement as directed at what he describes as *existing capacities*, in order to avoid reference to notions of moral normalcy, and the problems associated with such accounts, there are some thinkers who take the opposite approach. As mentioned above, enhancement may also be defined as referring to any improvement above normal or species-typical (human) levels of functioning. Defining enhancement in this way is associated with the somewhat contested distinction between treatment and enhancement that has been discussed at length in the general enhancement literature (Buchanan et al, 2009; Bostrom &

Roache, 2008; Daniels, 2000; Harris, 2007; Holtug, 1998; Juengst, 1998; Savulescu et al., 2011). According to this view any intervention that targets human functioning, dispositions or capabilities that are below species-typical levels in order to elevate them to 'normal' levels, would be regarded as the treatment of an impaired condition; whereas any elevations above species-typical levels, would be regarded as enhancements. This distinction is a useful one; but, as mentioned above, the problem of defining what would constitute normal species functioning is not a straightforward endeavour, particularly in the case of moral functioning. This is because functioning tends to occur on a spectrum, thus, there is no clear division between when treatment of impaired functioning would become enhancement of normal functioning.

This approach is nevertheless useful in responding to arguments such as those of Persson and Savulescu who argue for moral bioenhancement as a means of improvement or correction of the flaws in human moral psychology (2012). This would amount to the view that what is required to secure the future of humanity from ultimate harm would be a collective *treatment* of humanity's moral shortcomings. However, as mentioned above, to take this approach as a means of justifying a programme of moral bioenhancement, it would be necessary to establish a level of moral functioning taken as optimal. It would also have to be established that this level of moral functioning would be high enough to safeguard against the kinds of risks of ultimate harm that Persson and Savulescu are concerned about. Beauchamp makes a similar point in arguing that "establishing threshold levels seems critical for a project of bioenhancement that starts from the premise of existing moral deficiencies" (2015:346).

These problems aside, reframing the debate as one that requires moral treatment rather than enhancement would be an interesting, and perhaps preferable approach for those who support the argument put forward by Persson and Savulescu. Firstly, there would potentially be more acceptance of a programme of treatment or therapy of deficiencies in moral functioning. This is evidenced by the fact that we currently have a variety of interventions, both biological and non-biological, that are used to treat psychological pathologies, such as personality disorders, which directly influence moral behaviour. Most of these interventions and treatments are accepted without controversy. Secondly, as pointed out by Casal, reframing the problem as one that requires treatment could assuage the fears of those who view the project of moral bioenhancement as posing a "risk of changing human nature beyond recognition and losing our moral compass" (2015:341).

Wiseman who defines moral enhancement in terms of hard and soft forms, which loosely correlates with the treatment/enhancement distinction, has also argued that a soft kind of moral enhancement used to bring individuals up to normal levels of functioning could be argued for more easily than the hard form (2014:48). As pointed out by Kahane and Savulescu, by enhancement, the supporters of moral bioenhancement are generally referring to *normal range human enhancement* (2015:133). While there are a number of supporters of radical transformation of human capabilities, particularly those who wish to create super-intelligent posthumans and radically extend human life, Kahane and Savulescu argue that "this focus on…*supranormal enhancement* can be an obstacle to clear thinking about the forms of biomedical enhancement that are almost certainly feasible, and in fact likely to be available soon" (2015:133). Discussions of radical enhancement taking place elsewhere in the enhancement literature suggest the idea that supporters of moral bioenhancement wish to introduce a "radical new element to our mental life" (Kahane & Savulescu, 2015:134), when in actual fact they are wishing to "modulate naturally existing substances and processes" (ibid.).

In terms of using the treatment/enhancement distinction to define moral bioenhancement, the foremost exponent of this approach is Nicholas Agar. He posits that moral therapy would be focused on "measures designed to boost responsiveness to ethical or moral reasons to levels properly considered normal for humans. Moral enhancement…[on the other hand, would have] the purpose of boosting responsiveness to ethical or moral reasons to levels beyond that considered normal for human beings" (Agar, 2010:73). Whilst Agar regards moral treatment as relatively unproblematic, he is not a supporter of moral bioenhancement which he regards as dangerous (2015a:343). Agar provides a number of arguments against moral bioenhancement – which will be discussed further in this chapter and chapter 3 – which he defines as: "the use of biomedical means, including pharmacological and genetic methods, to increase the moral value of our actions or characters" (2015b:37).

As mentioned above, DeGrazia specifically uses the notion of "*enhancement as improvement"* of capacities in his definition of moral enhancement, choosing to avoid references to "*enhancement beyond moral norms"* (2014:369). While Agar endorses a "conceptual pluralism" (2014:369) regarding the existence of a variety of ways of defining moral bioenhancement, he criticises DeGrazia's interpretation in this context due to the implications it has for obfuscating particular arguments against moral bioenhancements. When what is described as moral bioenhancement is defined in terms of taking human moral capacities beyond moral norms then certain problems, such

as the impact upon human freedom of such enhancements, emerge that are able to be more easily side-stepped when it is defined in terms of improvement. Agar argues that DeGrazia is possibly aware of this and it may be why he chooses to define moral bioenhancement simply in terms of improvement. According to Agar, it is also why DeGrazia's argument that moral bioenhancement would not compromise freedom succeeds. If Agar is correct then this would be an example of the claim that the way in which moral bioenhancement is defined is reflective of an underlying agenda and is thus, not a neutral endeavour.

2.2.4 Other definitional distinctions

Before concluding this section, some of the useful insights from the clarificatory taxonomy provided by Raus et al must be mentioned. Raus et al.'s taxonomy is aimed at organising and categorising the various definitions of moral bioenhancement (2014). One of the ways in which definitions may be categorised is in terms of broad or specific *interventions* (Raus et al., 2014:265). Under this understanding, broad interventions would simply refer to defining enhancement in such a way that it could refer to both more invasive biological enhancements that act on the body in some way, as well as more traditional mechanisms of enhancement, such as moral education and socialisation. Specific interventions would refer to the former only, and would be associated with stronger ethical concerns. This is a similar point to the one that I made above regarding DeGrazia's definition of moral bioenhancement. In other words, utilising a broad definition would serve to minimise the appearance of differences between traditional and biological enhancement and thus garner support for the latter based upon acceptance of the former.

Another interesting way in which one may define moral bioenhancement would be to distinguish between active and passive enhancements (Raus et al. 2014:270). This point is related to the argument, mentioned above, made by Focquaert and Schermer (2015a) regarding the ethical relevance of the means that one utilises to achieve enhancement. The legitimacy of this distinction has major implications for attempts to conflate the distinction between traditional and biological moral enhancement. According to this distinction, moral bioenhancements would be regarded as passive enhancements in that they require minimal effort on the part of the individual and would therefore be regarded as more easily able to bypass the reasoning faculties of individuals. In this light, moral bioenhancements could be regarded as a form of behaviour control, and, would thus be regarded as controversial. Traditional moral enhancement, on the other hand, would be viewed as an active enhancement as the argument would be that the individual is integrally involved in

directing, contributing and questioning the process of moral education (Raus et al. 2014:270). I will discuss this important distinction in more detail in chapter 4

2.3 The problem of moral content

As mentioned above, the second area of focus that is relevant for the research focus of this dissertation is the primary source of the problem of competing definitions. This problem is one that is as old as the philosophical endeavour that seeks to answer the question regarding what makes a person moral, or what can be considered a morally good life. The problem of moral content is thus concerned, on one level, with the question of what should be the target of moral bioenhancement – should it target behaviour; motives, emotions or feelings; the reasoning processes that bring us to moral conclusions or a combination of all of these? In other words, which of the targets, or combinations of targets discussed in section 2.2.2 should be the legitimate focus of moral bioenhancement? Furthermore, there is the deeper issue regarding how the answer given to this question is justified.

This area of investigation is not only conceptually or theoretically interesting; it also has practical implications regarding the scientific feasibility of moral bioenhancement. For moral bioenhancement to be viable, consensus must be reached regarding what it is we should enhance and it must, of course, be scientifically feasible. In other words, regarding the latter point, it must be shown that the target of moral bioenhancement has biological origins or is susceptible to biological moderation. The issue of justification is a more challenging one as it requires addressing seemingly insoluble meta-ethical disputes. Furthermore, moral pluralism – the view that there are a variety of distinct and equally legitimate values or perspectives regarding 'the good' which cannot be conflated and which may conflict with one another – is such, that reaching consensus here may not be possible. As mentioned above, a vast proportion of the disagreements regarding what should be the target of moral bioenhancement are informed by the dispute between Hume and Kant regarding the basis of morality. This dispute, that I have coined the emotion/reason dichotomy, investigates questions such as the extent to which morality is an emotionally driven phenomenon, or, a product of our ability to utilise our reasoning to transcend our emotions. In addition, other normative disputes between moral theories inform what is taken as the target of moral bioenhancement and its justification. The moral bioenhancement debate, comprises supporters such as Persson and Savulescu, Douglas, Walker and DeGrazia and opponents such as Harris, and a variety of other thinkers, is essentially one between a consequentialist justification of moral

bioenhancement and a non-consequentialist response[5]. It is difficult to see how such a debate could be resolved.

2.3.1 Moral motivation – the view of the supporters

As mentioned in the first chapter, Persson & Savulescu launch their argument for moral bioenhancement on the basis that there is a lag between human moral psychology and the urgent and existence-threatening challenges humanity faces in the 21st century. They argue that our moral psychology evolved in very different circumstances to deal with existence in small communities and is therefore ill-equipped to address the variety of challenges we now face living in vast societies. The solution to the various flaws in our moral psychology, that Persson & Savulescu discuss at length, is to biologically enhance our altruism and sense of justice (2012). The biological basis of these dispositions and the way in which they could be enhanced will be discussed in the following section.

Persson & Savulescu argue that while moral progress has been made in terms of humanity acquiring greater knowledge of what is good[6]; they do not agree with the commonly accepted Socratic posit that simply knowing the good implies doing the good (2008:168)[7]. They argue that there is a gap between the two that is the territory of moral motivation or "moral will" (Persson & Savulescu, 2010:666). It is this "motivational insufficiency" (Persson & Savulescu, 2013:129) that must be strengthened and enhanced. Thus, Persson & Savulescu suggest targeting particular moral dispositions – altruism and our sense of justice – because they believe that enhancing these dispositions will be the best means of strengthening our moral motivation and of closing the gap between knowing the good and acting upon it.

---

[5] These arguments are non-consequentialist or Kantian in nature due to the fact that they emphasise the intrinsic value of phenomena that they argue would be threatened by moral bioenhancement. In particular, such arguments emphasis the value of moral autonomy as the foundation of morality, and the primacy of the means taken to reach moral conclusions or outcomes.

[6] An example of moral progress that Persson & Savulescu mention is "the doctrine of equal worth of all human beings" (2010:667) that is now espoused by most democratic nations. However, they argue that this doctrine has not been internalised sufficiently for it to compel us to address pressing issues such as global inequality (ibid.).

[7] To illustrate this point, Persson & Savulescu discuss the phenomenon of prejudices such as racism, xenophobia and homophobia (2013:129). Harris has argued that such prejudices are "likely to be based upon false beliefs about those racial or sexual groups; and, or, an inability to see why it might be a problem to generalize recklessly from particular cases" (2011:105). In other words, he argues that such prejudices have "cognitive content" and could therefore be dispelled through "a combination of rationality and education" (Harris, 2011:105). In response, Persson and Savulescu argue that this approach would only address part of the problem: "the mere realization that racism is false is not enough to wash away all xenophobic reactions in our nature" (2008:168). They argue that racism is part of a host of xenophobic reactions "that evolved to detect coalitional alliances" and seems to be the product of "computational processes that appear to be both automatic and mandatory" (Persson & Savulescu, 2008:168).

It must be noted that Persson and Savulescu do not deny that individuals may possess a strong sense of justice and altruism; rather, what concerns them is that these dispositions are primarily displayed towards *in-groups* – those we consider to be connected to us in some way - rather than *out-groups* – those perceived as strangers (2013:129). They argue that "to be morally good involves not just knowing what is good, but being so strongly motivated to do it that this overpowers selfish, nepotistic, xenophobic, etc. biases and impulses" (Persson & Savulescu, 2013:130). Furthermore, their aim in this regard is not radical moral bioenhancement, rather it is that "the moral motivation of those of us who are less morally motivated be increased so that it becomes as strong as the moral motivation of those of us who are by nature mostly morally motivated" (Persson & Savulescu, 2012:113). As discussed in section 2.2.2, Douglas also regards motives as the target of moral bioenhancement. Thus, he concurs with Persson and Savulescu that morality lies in closing the gap between knowing what is right and doing what is right (Douglas, 2013:162). Whilst he isn't necessarily a supporter of moral bioenhancement due to its potential freedom-subverting consequences, Rakić argues that "the discrepancy between what we do and what we believe it right to do might be the greatest predicament of our existence as moral beings" (2012:120)[8]. Thus, he concurs with Persson & Savulescu that it is "motivation rather than cognition that is at the heart of the matter" (Rakić, 2014:248).

Of course, defining moral bioenhancement simply in terms of moral motivation or increasing the likelihood of doing "the right thing and for the right reason(s)" (Savulescu, Douglas & Persson, 2014:95) glosses over the fact that there isn't consensus regarding "what accounts of right action and right motivation are correct" (ibid.). In other words, depending upon the moral theory one supports, conceptions will differ in this regard. For a Kantian, it is not simply having good motives in a general sense that confers moral worth upon one's actions; rather, it is acting from the motive of duty (Kant, 2002), where one utilises the dictates of reasoning to work out the nature of one's duty in accordance with the categorical imperative. One could state this in an even stronger manner by positing that, according to a Kantian, truly employing one's rational capacities requires that one must transcend emotional interference as "emotions lie outside of the boundaries of the will" (Douglas, 2008:232).

Consequentialists, and in particular, Utilitarians, on the other hand, would see the correct motives as those which ensure the best outcome, all things considered. Furthermore, for a Humean, we are

[8]It is interesting to note that this intuition is a long-standing one. In the Christian bible St Paul makes a similar posit when he argues that there is a continual conflict within 'man' who despite wanting to do what is good, is seemingly compelled to rather do what he knows is bad (Romans, 7:18-20).

constituted by our passions, therefore viewing motives – and morality in general – in purely rational terms, devoid of emotional content, is flawed. Hume states that:

> Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be deriv'd from reason; and that because reason alone, as we have already prov'd, can never have any such influence. Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason (Hume, 1978:239).

2.3.2 The emotion/ reason dichotomy

Due to the afore-mentioned disputes, the debate between Persson & Savulescu and Douglas, on the one hand, and thinkers such as Harris, Agar and Sparrow, on the other hand, is often framed as one between the role of emotions versus reasoning in morality[9]. The reason for this is twofold. Firstly, there is the view that enhancing the dispositions of altruism and our sense of justice, as a means of strengthening our moral motivation, can be interpreted as intensifying an emotion or feeling that compels us to act in particular manner. Thus, altruism and a sense of justice are viewed as decidedly affective or emotional in content. In addition, moral bioenhancement, according to Douglas, should consist of biomedical interventions that ameliorate *counter-moral emotions* that interfere with moral motivation. He explicitly states that morality can be enhanced cognitively or non-cognitively with the latter referring to moral enhancement "achieved through (a) modulating emotions, and (b) doing so directly, that is, not by improving (increasing the accuracy of) cognition" (Douglas, 2013:162). It is this direct modulation of emotions that is seen as problematic by thinkers such as Harris who argue that such modulations would bypass our reasoning, and thus, that they would not constitute an enhancement but would simply be a form of behaviour modification or control (2016:99).

The second reason that the debate can be framed as one between enhancing or attenuating emotions versus moral reasoning is due to Persson and Savulescu's argument regarding cognitive enhancement. Persson & Savulescu have argued extensively against the dangers of cognitive enhancement, if it is not accompanied by moral bioenhancement, implying that the two domains are distinct from each other (2008)[10]. Carter and Gordon disagree with this view and provide an

[9] Baertschi, argues that in this regard the moral bioenhancement debate is at heart a meta-ethical one (2014:66). This is perhaps why it is such a fascinating area of discussion, but also the reason why it is seemingly irreconcilable. Baertschi argues that Douglas and Persson and Savulescu are *sentimentalists* and utilitarians, thus, they focus on results; whereas Harris and many other opponents of moral bioenhancement are rationalists and for them "intentionality and consciousness are at the core of morality" (2014:64).

[10] They argue that cognitive enhancement in the power of an immoral individual, or groups of individuals, may intensify the risk of ultimate harm as the knowledge required to create both biological and nuclear weapons of mass destruction will be easier to acquire for someone with heightened intellectual abilities.

argument for their claim that "just as there is a moral dimension to cognitive flourishing, there is a cognitive dimension to moral flourishing" (2013:158). Therefore, they argue that a true programme of moral bioenhancement cannot disregard the important cognitive component. Harris, being a rationalist, on the other hand, is a supporter of cognitive enhancement, but not moral bioenhancement – as construed by Persson and Savulescu.

Persson and Savulescu do not, however, disregard the importance of processes of reasoning in addressing incorrect moral beliefs. While they have specifically opted for a narrow interpretation of moral bioenhancement, they do clearly state that it may be defined in a wider sense, as moral enhancement, which would include cognitive enhancement, either through traditional or biomedical mechanisms (2015a:349). Their argument, however, is that cognitive enhancement, in whichever form, will not be sufficient alone to enhance moral motivation to the extent that it will be able to address the threats to existence they outline. They explicitly state that:

> although moral bioenhancement of the powers of reason and intellect is not part of what we mean by moral bioenhancement, this doesn't imply that these powers can be bypassed or rendered superfluous if we want to achieve moral improvement. It only implies that being moral isn't *solely* or *exclusively* a matter of the operation of intellectual or rational powers, but this is something that few would deny (Persson & Savulescu, 2015a:350).

As mentioned above, another distinction that is often made in the literature, which closely resembles the emotion/reason dichotomy, is between moral bioenhancement as behavioural control and what is framed as 'true' moral enhancement which would be akin to acting morally for the right reasons. Those who view moral bioenhancement as behaviour modification generally hold this view because they interpret the supporters of moral bioenhancement as targeting emotions[11]. Simkulet argues that framing moral bioenhancement in terms of behaviour or "detectable modifications of one's moral conduct…fails to distinguish moral enhancement from moral compulsion" (2012:17). Jotterand argues that interventions can produce "changes in mood, affect and behaviour. However these techniques do not focus on morality…[rather, they are altering] how people react to situations that implicate a particular moral stance" (2014:2).

Persson & Savulescu's conception of moral bioenhancement is often criticised on these grounds. This may be due to the emphasis they place on the enhancement of motivation as the best means of closing the gap between knowing and doing the good. In other words, they focus on

[11] Jebari, however, makes the point that emotional enhancement is not one and the same thing as behavioural 'enhancement' (2014:255). Rather, emotional enhancements alter how we perceive particular behaviours. For example, after being emotionally enhanced I may be repulsed by aggressive or violent behaviour which disinclines me to act in such a manner (Jebari, 2014:255). This would accord with Douglas' view of moral bioenhancement as the attenuation of counter-moral emotions.

enhancements that would compel us to act, or make it more likely that we will behave in a particular way. Their view is that while reasons and moral knowledge are important, in the absence of changes in action or behaviour, they do not constitute moral improvement. However, they take exception to the portrayal of their moral bioenhancement aims as targeting only emotions and behaviour. Sparrow, in particular, makes this charge and argues that for Persson & Savulescu, modifying behaviour and feelings amount to moral enhancement (2014a:24). However, Persson and Savulescu point out that they are not targeting feelings in general. Rather, their target is the specific dispositions of altruism and a sense of justice. Individuals would still, presumably, act according to particular reasons, however, their motivation to act upon these reasons would be strengthened. Nevertheless, the conception of moral bioenhancement as behaviour modification fuels the argument that moral bioenhancement will somehow impact upon, or destroy, our moral autonomy. Harris is the primary proponent of this view, arguing that morality presupposes our "freedom to fall" (2011:104). In other words, an act can only be truly moral if it is made in the face of competing possibilities to have acted otherwise; even wrongfully. This concern is, of course, the main focus of my dissertation and will thus be discussed extensively in chapters 4 and 5.

2.3.3 The 'problem' of moral pluralism

Regarding the issue of competing moral justifications, Sparrow has, for example, argued that Persson and Savulescu's programme of moral bioenhancement attempts to impose a particular interpretation of 'the good' and is thus at odds with "an egalitarian commitment to liberal neutrality" (Persson & Savulescu, 2014b:41) which respects an ethos of moral pluralism (Sparrow, 2014b:22). Brooks has similar concerns regarding the possibility that moral bioenhancement will undermine moral pluralism (2012:29). However, to this, Persson and Savulescu respond that while liberal neutrality recognises that individuals possess the freedom to conduct themselves as they see fit, this freedom is not absolute. It has restrictions regarding its impact upon others. Liberal neutrality should also not be confused with moral relativism, they argue. In other words, a respect for a diversity of moral perspectives does not imply that moral perspectives are simply a product of social or cultural context and that all of them are, therefore, correct in their own way. There are some moral perspectives that are regarded as more universally held than others. The move towards the recognition of the doctrine of equal moral worth of all human beings and the way in which research ethics is standardised and regulated by internationally recognised agreements, is evidence

that there is far more consensus regarding ethics, and 'the good' than is sometimes implied in moral bioenhancement discussions (Persson & Savulescu, 2014b:41)[12].

Whilst a number of thinkers argue that moral bioenhancement will not, or may not, be feasible due to a lack of consensus regarding the content of morality (Beck, 2015; De-Melo Martin & Salles, 2015, Schaeffer, 2011), Savulescu, Douglas and Persson argue that despite moral pluralism, there are acts that are almost universally regarded as right and wrong. The belief that "it is wrong to kill an innocent person in non-extra-ordinary circumstances" (Savulescu, Douglas, & Persson, 2014:95), is one such act that is endorsed by virtually all moral theories[13]. Pacholczyk also makes this point and argues that killing indiscriminately for pleasure, lying and breaking promises, without legitimate reasons for doing so, are generally viewed as wrong; whereas "concern and respect for other moral agents" (2011:174) are viewed favourably. DeGrazia suggests that in seeking to establish what would be considered an enhancement we should "stick to improvements that represent *points of overlapping consensus among competing, reasonable moral perspectives*" (2014:364). He posits that:

> leading contemporary progressive and conservative visions – ranging from socialism to welfare-state capitalism to moral conservatism to libertarianism – count as reasonable whereas neo-Nazism, apartheid and the Taliban's worldview (at least as regards women's status) do not. Consequentialist, deontological, virtue-based and feminist views that accord persons some sort of moral equality qualify as reasonable; Nietzschean elitism according to which only the most powerful and creative are worthy does not (DeGrazia, 2014:364).

DeGrazia also provides an extensive list of "moral defects" that most reasonable individuals would concur with. DeGrazia views moral improvement as characterised by improvements in motivation, insight and behaviour. Thus, his list of specific examples of moral defects consists of various failures in moral motivation and insight (DeGrazia, 2014:364). Regarding the problem of moral pluralism, DeGrazia points out that it is not only the project of moral bioenhancement that must address this difficulty; any form of moral education or socialization must take a particular stance on the content of moral norms (2014:363). Persson & Savulescu also address this issue and argue that the view that there is a level of ethical consensus is supported by the fact that there is considerable overlap in the way in which children are socialised and morally educated (2015a:349).

[12] Harris supports this view and argues for "the generic character of the good" (2016:16) at length in his book: *How to be Good* (2016). Pacholczyk also discusses this point and posits that disagreements regarding moral content are often exaggerated and prevalently used to refute moral realism and support a thesis of "the metaphysics of morals" (2011:174).

[13] Of course, the immediate response to this could be that while this may be true, the differences lie in conceptions of innocence. For example, in some societies, killing a woman who has committed adultery would not be regarded as killing an innocent person.

In this regard, Kahane and Savulescu argue that "it is hardly controversial that some minimal level of altruism is desirable within any morality, deontological, utilitarian or other" (2015:142).

In terms of the moral theory they espouse, Persson and Savulescu's argument for moral bioenhancement is justified in terms of the good it will achieve, and is thus, distinctly consequentialist, and, more specifically, utilitarian. Altruism and a sense of justice are not offered as candidates for moral bioenhancement due to their purported intrinsic value; rather, they are supported by consequentialist justifications. Persson and Savulescu posit that utilitarianism espouses the view that "the fundamental moral motivation is nothing but universal altruism or benevolence" (2015a:348). However, because utilitarianism is uniformly criticised as leading to outcomes that may be at odds with the requirements of justice; Persson and Savulescu modify their form of utilitarianism into a "two principle moral theory" (2015a:349) with the inclusion of the moral disposition of a sense of justice. They regard the flaws of our common-sense morality, described in the first chapter, as characterised by "deontological features" (Persson & Savulescu, 2015a:349). As most individuals struggle to abide by even this flawed common-sense morality, the argument is that moral bioenhancement is necessary in order to adhere to the more demanding "proposed consequentialist extension or revision of it" (Persson & Savulescu, 2015a:349).

2.3.4 Harris' position

There are a number of thinkers who oppose moral bioenhancement on the grounds that it will either circumvent or overshadow the cognitive component of morality (Agar, 2015a, 2015b; Harris, 2011, 2016; Sparrow, 2014a, 2014b). Harris, for example, has always supported bioenhancement, therefore, his opposition to moral bioenhancement is interesting. However, he is not opposed to moral bioenhancement *tout court*. He supports moral enhancement via cognitive enhancement, where the latter may include traditional mechanisms such as socialisation and moral education or cognitive bioenhancement (Harris, 2011:102). This is because, for Harris, "ethical expertise is not 'being better at being good', rather it is being better at knowing the good and understanding what is likely to conduce to the good" (2011:104). Harris argues that a moral bioenhancement that targets emotions amounts to doing "ethics with [one's] gut" (2013a:288). Implying that emotional responses are adequate in addressing complex moral problems "is like believing the gut is an organ of thought" (Harris, 2013a:288). Drawing upon the work of Ronald Dworkin, Harris argues that moral judgements are respected due to the fact that they are a product of deliberation and consideration (Dworkin, 1977). Furthermore, moral judgements are subject to particular requirements, they cannot include *disqualifying features* such as "gut reactions, and instinctive or

automatic responses…[and must] be distinguishable from prejudices, arbitrary preferences, personal tastes, arguments or conclusions based on manifest self-interest or partiality, or arising from a personal emotional response" (Harris, 2013a:289).

While Harris admits that emotions play some part in moral decisions, he argues that "it must be reasoning that pulls in the direction of morality" (Chan & Harris, 2011:130). Our reason must act as a watchdog on the emotions because we cannot always know whether our feelings are informed by the above-mentioned disqualifying features. In other words, Harris argues that if we define what we take to be 'the good' to involve "feeling the right way, how do we know that we are feeling the right way" (2016:113). Furthermore, the problem with viewing morality as consisting of feelings alone is that the test for ascertaining the legitimacy of one's feelings will generally involve "a second consultation of one's feelings" (ibid). This, Harris argues, is akin to Wittgenstein's example of purchasing an additional, but identical, copy of a newspaper to validate what one has read in the first newspaper (Wittgenstein, 2001:79). As Wittgenstein argues, "justification consists in appealing to something independent" (in Harris, 2016: 126). This independent contribution is the role played by moral reasoning and judgment. For Harris, if a moral enhancement results in a change in behaviour without the input of cognition then this is not a true moral enhancement (2013b:172)[14].

Implicit in this respect for a rationalist account of morality is the view that acts that are the product of deliberation have more moral worth. Thus, improvements in moral conduct which are a product of moral bioenhancement would be seen as less morally worthy due to the perception that the individual in question has not made an active contribution. In other words, as Baertschi points out, "morality is like climbing. Climbing a mountain is not the same as reaching its top. It depends on the means used" (2014:64). Reaching the top of a mountain via helicopter would not be viewed in the same light as having climbed it. Jebari points out that this view is very similar to the Aristotelian view that virtues are manifested through habitual action (2014:259). Individuals become virtuous through the effort of acting in a virtuous manner. Supporters of such a view would argue that moral bioenhancement is akin to using blood doping to win a race and thus amounts to cheating because it isn't accompanied by any effort on the part of the individual. However, Jebari argues that the analogy doesn't hold. A race is subject to particular rules and ways of winning that are either

[14] As will be discussed further in section 2.4, Agar opposes moral bioenhancement for similar reasons to Harris, arguing that moral bioenhancement will threaten the balance between "cognitive, emotional and motivational subcapacities" (Agar, 2015:343), as it will involve the "bypassing of reason" (ibid.:345).

viewed as legitimate or not; whereas morality is not a game or a competition. If I am more moral it doesn't disadvantage others; quite the contrary, in fact (Jebari, 2014:259).

The connection between moral worth and effort is an extremely interesting one. On the one hand, morally worthy acts made in the face of temptations to have done otherwise are greatly respected, as pointed out by Harris (2011). On the other hand, Sorensen has correctly pointed out that we are equally impressed by morally worthy acts that are easily and readily performed without question on the part of the individual performing them (2014:282). In other words, we value, and view as morally worthy, both acts that require effort, and those that occur in a seemingly effortless manner. This matter will be discussed further in chapter 4.

2.3.5 The complexity of morality

Of course, it could be pointed out that viewing the moral bioenhancement debate in such polarized terms is an oversimplification of morality. Many thinkers posit that morality is not purely emotive or rational; rather, both play an equal role (Baertschi, 2014, Bublitz, 2016; Focquaert & Schermer, 2015a, Pacholczyk, 2011, Jotterand, 2014)[15]. In this regard, there are more nuanced understandings of morality in the literature. Focquaert and Schermer – drawing upon Fischer and Ravizza (1998) – interpret morality or *being moral* as being characterised by *moral responsibility* where having the "capacity for moral responsibility implies that an individual has the ability for *reason-responsive* behaviour, comprising both a 'receptivity to reasons' and a 'reactivity to reasons'" (2015a:141). Whilst receptivity to reasons tracks cognition and reactivity to reasons tracks emotions, this interpretation emphasises the interplay between the two. Pacholczyk also describes the complexity of the moral sphere, referring to the multiple components that comprise it; namely, "the ability to make moral judgements, to be motivated by moral reasons, acting according to our moral beliefs, the ability to reflect on and critically analyse moral beliefs, and so on…[and the fact that these capacities rest on both] affective and cognitive capacities" (2011:170). Another important point she makes is that "moral behaviour encompasses a range of behaviours and includes refraining from doing what is wrong, doing what one ought to and doing good things beyond one's obligations" (Pacholczyk, 2011:170). The last behaviour is an important one as it captures the ethos of morality; namely that it involves doing what one ought to do even when it isn't personally beneficial; simply because it is the right thing to do.

[15] While the primary proponents and opponents in the debate (Persson & Savulescu, Douglas and Harris) would agree that that morality is constitutive of both elements, their arguments imply that they, nevertheless, regard one component as having primacy over the other.

However, it is not only the issue of moral content, regarding the affective and cognitive content of morality, that is prevalent in the literature. The other issue, mentioned above, is described by Chan and Harris as "one of the perennial problems of moral philosophy: how we measure morality, whether it is about ends and the ultimate outcomes of action or about means and reasons to act, the drivers underlying the action" (2011:131). Referring to Persson & Savulescu's explicit justification of moral bioenhancement with the utilitarian aim of the avoidance of ultimate harm, Harris argues that "to prioritise the avoidance of so-called direct harms is neither the action of a good conscience, nor of a good consequentialist" (2013c:119). Elsewhere, he reiterates that morality is not simply about the avoidance of harm (Chan & Harris, 2011:131).

Agar also discusses the problem, arguing that while there is consensus between different ethical theories that it would be good to reduce phenomena such as murder or famine, their reasons as to why this would be good to do, and how it should be done, differ (2010:74). These differences manifest themselves keenly in the moral bioenhancement debate as it is founded upon the aim to improve the world in some way. Furthermore, the ethical theories draw their views of morality from different interpretations of human ontology: utilitarians give primacy to the "capacity for suffering and enjoyment…[while] Kantians prize practical rationality" (Agar, 2010:75).

Hauskeller makes an interesting point regarding disagreements among moral theories. He argues that supporters of moral bioenhancement, such as Persson and Savulescu, are not arguing for moral bioenhancement on the basis that making people more moral is an intrinsic good, "morality [for them] is not the end. It is the means" (Hauskeller, 2015:290). In other words, Persson and Savulescu have opted for moral bioenhancement purely because they see it as the most effective means of addressing urgent global problems. He argues that recognising this vital difference "between morality being pursued as an end and as a means is likely to deflate the heated debate on whether or not people can, and should, be morally enhanced, and make the idea more palatable to critics" (Hauskeller, 2015:290). One could then engage with Persson and Savulescu's argument on a different level and assess it in terms of speculations regarding its effectiveness in achieving its aims, rather than in terms of whether or not their conception of moral bioenhancement represents a true moral enhancement.

2.3.5 Virtue ethics as an alternative

Whilst the debate is primarily dominated by utilitarian justifications for moral bioenhancement and non-consequentialist or deontological responses, virtue ethics has also been utilised by a number of thinkers in the moral bioenhancement debate as an alternative theory (Jotterand, 2011; Hughes, 2013, 2015; Fröding, 2011; Walker, 2009). In this regard, both Hughes and Jotterand argue that the moral bioenhancement debate is excessively dominated by the Kantian/Humean polarization between reason versus emotions as the basis of morality, whereas virtue ethics offers a more integrative framework (Hughes, 2015:86; Jotterand, 2011:6).

Jotterand posits that both reason and emotions are indispensable to morality; however, he argues that even if we are able to modify emotions, thereby influencing behaviour, we will still be in need of moral content "for example, norms or values to guide one's behavioural response" (2011:6). Jotterand distinguishes between "*character traits* and *having character*" (2011:8), where the former is behavioural and not necessarily moral in nature. Having character, on the other hand, consists of "agency (reasons, motives, intentions) and action" (Jotterand, 2011:8). Jotterand argues that moral bioenhancement, as presented by its supporters, would modify character traits but not character itself. Virtue ethics could therefore be helpful as Jotterand posits that "it takes into account the fullness of human experience, i.e., emotional motivational and rational dimensions" (2011:6). The gist of his argument is that using a virtue ethics framework to improve human morality would require enhancement of both rationality (cognition) and the relevant emotional dispositions. Fröding, on the other hand, argues that cognitive bioenhancement in conjunction with the insights of virtue ethics could be justified on the grounds that it would lead to a good life characterised by flourishing (2011:223).

Hughes uses virtue ethics in a more concrete manner than Jotterand. He draws upon empirical research in the fields of psychiatry, neuroscience and moral psychology to identify four potential candidates for biomedical virtue enhancement: *self-control*, *niceness*, *intelligence* and *positivity* which are all neurobiologically mediated (Hughes, 2015:89). Hughes posits that these virtues he has identified correspond roughly with "the four cardinal virtues of Plato and Aquinas – temperance, justice, prudence and courage" (2015:29). Elsewhere, Hughes has discussed the connection between Buddhist ethics and virtue ethics, arguing that the former may assist in providing candidates for virtue enhancement (2013:28). Walker, a staunch advocate of moral bioenhancement, also utilises a virtue ethics framework to identify personality traits that are genetically influenced as potential candidates for enhancement. In his description of a hypothetical

inter-disciplinary Genetic Virtue Programme, He focuses on "truthfulness, justice and caring for others" (2009:15) as plausible candidates for bioenhancement and argues that they would be "the best mechanisms not to make persons virtuous but to make them better equipped to learn how to be virtuous" (Walker, 2009:28)[16].

2.4 The problem of the science of moral bioenhancement

As mentioned above, an investigation of the scientific feasibility of moral bioenhancement is, of course, related to the first two problems that I have identified and discussed. This is because, as mentioned in section 2.1, biological enhancement of our morality presupposes that we have pinpointed and reached consensus regarding what it is we are actually aiming to enhance. Whilst the moral bioenhancement debate is, at this point, primarily theoretical, there has nevertheless been discussion of the biomedical interventions that could be utilised. The arguments provided by the proponents of moral bioenhancement tend to be more optimistic regarding the possibilities offered by biomedical science, whereas the contributions from a number of biologists and neuroscientists are somewhat more cautious.

Scientific discussions can therefore be categorised according to, on the one hand, those who are enthusiastic regarding the scientific possibilities of moral bioenhancement (Walker, 2009; DeGrazia, 2014) and those who are optimistic, but cautious (Persson & Savulescu, 2012; Douglas, 2008; Spence, 2008; Pacholczyk, 2011). On the other hand, there are those who, for various reasons, see moral bioenhancement as unfeasible. This group would include those who definitively think it will not be possible (De Melo-Martin & Salles, 2015) and those who think it will most likely not be possible (Crockett, 2014; Sparrow, 2014). One of the reasons given by members of this second group is that it would be impossible – or inadvisable, even if possible - due to the complexity of human moral psychology (Zarpentine, 2013; Young & Duncan, 2012; Lechner, 2014; Bronstein, 2010; Andreadis, 2010; Arnhart, 2010; 2013; Sprinkle, 2010) Those who hold this view tend to argue that the side effects, risks or negative consequences will be too great (Harris, 2011; Clausen, 2010; Barilan, 2015; Jones), or, that moral bioenhancement will morally worsen us (Agar, 2015a; Chan & Harris, 2011, Hubbeling, 2009).

[16] There are a number of criticisms of such virtue ethics approaches and addressing this matter is beyond the scope of my dissertation; however, it is interesting to note that this approach has been criticised by Harris (2013c). With its focus on the character and dispositions of individuals, as well as the way in which virtues are manifested in habitual action, Harris seems to agree with the long-held belief that the virtues, and thus virtue ethics itself, are affective in content. Harris is not a utilitarian thinker; however, he posits that a focus exclusively on virtues – which he seems in this instance to equate with "moral emotions, or states of mind, or intentions" (2013c:119) – without consideration of consequences is dangerous.

2.4.1 Supporters of the scientific feasibility of moral bioenhancement

As most of the discussion in the moral bioenhancement debate has engaged with the various arguments made by Persson and Savulescu (2008; 2010; 2011; 2012; 2013; 2014a; 2014b; 2015a; 2015b), it seems appropriate to commence with their interpretation of the scientific considerations. As mentioned above, and discussed in chapter 1, Persson and Savulescu provide a substantial account of the failings of human moral psychology, which equipped us, through evolution, to deal with vastly different concerns to those faced in contemporary society. In this regard, they argue that human moral psychology is not up to the task of addressing the urgent global problems that characterise life in the twenty first century, particularly the danger of ultimate harm. To address these problems, they argue that what is needed is to enhance particular moral dispositions, namely altruism and our sense of justice. These dispositions, they argue, are "malleable by biomedical and genetic means" (2008:168) and are "central moral dispositions because they motivate us to act in accordance with plausible basic moral principles" (Persson & Savulescu, 2015a:348). They see altruism as associated with both empathy "in the sense of a capacity to imagine from the inside what it would be like to be another conscious subject, and…[benevolence in terms of a] sympathetic concern about the well-being of this subject for its own sake" (Persson & Savulescu,2015a:348].

They admit that justice is difficult to define clearly as a philosophical concept, and, that one cannot offer substantive claims regarding its nature without providing protracted supporting arguments (Persson & Savulescu, 2015a:348). Therefore, they explain their understanding of this notion through descriptions of its evolutionary origins. Their evidence for the biological origins of both suggested dispositions is drawn from the field of evolutionary biology and the research indicating that we find similar dispositions in animals with whom we share evolutionary origins. Persson and Savulescu argue that these dispositions played a role in human survival and evolution. The more sophisticated sense of justice that is characteristic of contemporary human psychology is argued by evolutionary biologists to have originated in the role that reciprocity would have played in survival. In other words, if an individual granted another individual a favour, the receiver would reciprocate out of gratitude; whereas if a favour wasn't returned then the common response would be anger and the tendency to not grant favours to that individual again in the future. A group or society in which such patterns of reciprocity and cooperation were established would be more cohesive and well-functioning, and therefore, more likely to prosper and survive. Persson and Savulescu posit that out of these primitive emotions, higher order dispositions, such as "remorse

and feelings of guilt…shame…pride...admiration and contempt…and forgiveness" (2008:169), arose. Altruism and this early sense of justice or "tit-for-tat" (Persson & Savulescu, 2008:169) response were interconnected in that this sense of justice would inform responses of anger to inappropriate reciprocity of altruistic acts, thereby encouraging future altruism and discouraging selfish behaviour.

Persson and Savulescu cite extensive sources, based on animal studies, that have documented the genetic foundations of altruism and the above-mentioned sense of justice (Sober & Sloan, 1998; De Waal, 2006). Their hypothesis of the biological foundations of these dispositions is also supported by twin studies in which remarkably similar responses to the ultimatum game occur with identical twins (Baron-Cohen, 2003:114; Wallace et al. 2007:15631-4)[17]. In ultimatum games that test the responses of identical twins, there are "striking correlation[s] between the average division with respect to both what they propose and what they are ready to accept as responders" (Persson & Savulescu, 2012:111). Fraternal twin studies do not deliver the same results (Wallace et al. 2007:15631-4).

In addition, Persson and Savulescu discuss the fact that, in general, it is accepted that a greater tendency for empathy, which they associate with altruism, is displayed by women (Baron-Cohen, 2003). Thus "if this psychological difference tracks gender, this is surely good evidence that it is biologically based" (Persson & Savulescu, 2013:13). Furthermore, if higher levels of empathy and altruism mitigate aggression and track gender then, in theory, Persson and Savulescu argue that "we could make men in general more moral by making them more like women by biomedical methods, or rather, more like the men who are more like women in respect of empathy and aggression" (Persson & Savulescu, 2013:130; Baron-Cohen, 2003:35). Persson & Savulescu point out that they are not arguing that socialisation and environmental influences do not play a significant role in the manifestation of such dispositions, just that a sizeable influence is biologically informed (2012:109). However, to those that deny any biological influences on our morality they argue that:

[17] Ultimatum games are an effective way of measuring the degree of strength of an individual's conception of justice. They generally involve two participants, a proposer who suggests a particular distribution of benefits and a responder who must either accept or reject the proposer's offer. If the offer is rejected then neither party receives anything. With human participants, the tendency is for offers that are unfair or disproportionately unequal, to be rejected, even if this means foregoing some benefits or receiving nothing. The view here would be that it is better to sacrifice some benefits than to support a situation in which another party acts in an unjust manner and unfairly benefits (Persson & Savulescu, 2012:35).

> every human behaviour, whatever its cause, is mediated by the final common pathway of the brain. Brain activity is just a series of electrical signals mediated by chemical reactions in the brain. One can modify any behaviour by modifying activity in the brain. That is basic neurobiology. So even if all moral behaviour were social in origin, one could still improve moral behaviour by moral bioenhancement just because the brain is the source of moral behaviour[18] (Persson & Savulescu, 2014b:42).

Of course, providing compelling argumentation and evidence for the biological basis, and thus, the potential malleability of what they take to be central moral dispositions is not sufficient. What remains to be seen is *how*, if at all, these dispositions could be biologically altered or enhanced. Persson and Savulescu admit that the science that would be required for this is currently in its infancy (Persson & Savulescu, 2008:172; 2010:667; 2013:130). However, there are currently a few rudimentary ways in which this could be done. One such way is through the administration of oxytocin, the hormone associated with "maternal care, pair bonding and other pro-social attitudes, like trust, sympathy and generosity" (Persson & Savulescu, 2012:118). Oxytocin is currently administered through a nasal spray but is also influenced by certain drugs of which oral contraceptive pills and glucocorticoids, used for the alleviation of asthma, are just two examples (ibid.). The effects of oxytocin administration on trust have been positively tested in cooperation games (Kosfeld at al., 2015). However, its effects on trust have been shown to "be limited to in-group members and exclude out-groups" (de Dreu et al., 2011 in Persson & Savulescu, 2012:119). In other words, oxytocin acts to strengthen trust, and thus, cooperation between individuals that share a perception of group-based similarities, and not between individuals who view themselves as having no commonly shared group characteristics.

In addition, Selective Serotonin Reuptake Inhibitors (SSRIs) – medications that are associated with blocking the reabsorption of the neurotransmitter serotonin in the brain – are associated with the tendency to cooperate and a decline in aggressive behaviour (Persson & Savulescu, 2008:172). These medications are generally prescribed and commonly taken for depression and anxiety. The effects of SSRIs have been measured in dictator games which involve the participation of a 'dictator' who has the freedom to decide how a particular sum of money shall be split between herself and another individual. It was found that when the SSRI, citalopram, was taken, it led to more fair divisions between participants than it did in the case of control groups (Tse & Bond, 2002). However, with increased levels of trust associated with the administration of SSRIs comes the concern that such individuals will be easier to manipulate. This concern is based on Crockett's

[18] Here they give an interesting example. It could be argued that something like the ability to read *could* be a wholly learned skill. But even if it has no biological basis, this does not mean that it cannot be improved through biomedical interventions, such as through the use of cognitive enhancers (Persson & Savulescu, 2014b:42).

findings that individuals are more likely to reject unfair offers in ultimatum games if they are low on tryptophan, a precursor of serotonin, which indicates that "SSRIs may make subjects easier to exploit by modulating their assessment of what counts as (unacceptably) unfair" (Crockett in Persson & Savulescu, 2012:120).

Despite the afore-mentioned concerns regarding oxytocin and serotonin, the important point that Persson & Savulescu are trying to make is that these experiments show that the brain modifications produced by the afore-mentioned drugs produce "moral consequences" (2012:121)[19]. Persson and Savulescu also discuss the possible biological basis of certain personality disorders such as antisocial personality disorder which is characterised by, among other things, a lack of cooperation and tendency to aggression. In addition, they mention research linking mutations on the X chromosomes which are connected to criminality, particularly when this mutation occurs in the context of particular environmental factors such as "social deprivation" (Persson & Savulescu, 2012:121).

Regarding the strength of Persson and Savulescu's argument for moral bioenhancement, they point out that their proposal is a *cautious* one which must be distinguished from more *confident* proposals regarding the likelihood of moral bioenhancement (2014b:39). They explain that a confident proposal would posit that "there *are* effective and safe biomedical means of moral enhancement waiting to be discovered, while a cautious proposal merely asserts that it's *possible* that there be such means" (own emphasis, Persson & Savulescu, 2014b:39). In addition, Persson and Savulescu draw attention to the fact that there is great variation in the distribution and levels of dispositions such as altruism and a sense of justice amongst individuals. They question why this "natural range" or "status quo" of distribution should be left unaltered. As mentioned in the previous section, they are not arguing for a radical form of moral bioenhancement (Persson & Savulescu, 2015a:349). There are many individuals in society who possess the kind of moral motivation required to address the type of problems facing humanity that Persson and Savulescu discuss. Thus, they believe that even modest bioenhancements would make a difference. Whilst the required scientific knowledge and technology may not yet be a possibility, they argue that it would be a good idea to address what they describe as "low hanging fruit, like the removal of tendencies to break the law and strengthening tendencies to commit acts of charity" (Persson & Savulescu, 2015a:349).

[19] Of course, Harris would respond to this by pointing out that this is evidence that what Persson and Savulescu call moral bioenhancement is, in fact, not true moral enhancement. Harris argues specifically that behaviour that has moral consequences is not necessarily "moral behaviour, any more than all behaviour that affects political outcomes is political behaviour" (2016:36).

DeGrazia is an enthusiastic supporter of the prospect of moral bioenhancement who has argued that there is nothing intrinsically wrong with moral bioenhancement. Furthermore, if it could be shown to be "safe, effective and universally available" (DeGrazia, 2014:361), he argues that it is something that we should go ahead with. While DeGrazia does not provide a detailed account of the scientific foundations of moral bioenhancement aims, he does provide an interesting list of current interventions that could be used for this end, as well as some future possibilities. He mentions research that has shown that the use of SSRIs reduces aggression (Crockett et al., 2010a), administering glucose assists in impulse control (Gaillot et al., 2007) and the beta-blocker Propranolol has been shown to reduce "unconscious racial bias" (DeGrazia, 2014:361, Terbeck et al. 2012.). More invasive inventions would include deep brain stimulation (DBS) to control aggression and neurofeedback to treat personality disorders associated with a lack of empathy and to boost exiting levels of empathy in general (DeGrazia, 2014:362). Genetic interventions would range from selecting embryos that possess a particular gene associated with a morally relevant trait such as altruism (Reuter et al., 2011), to avoiding selection of those embryos possessing genes associated with counter-moral traits[20] (Eley et al, 1999). This possibility is decidedly speculative, however, as it has not yet been established that there are genes associated with morally relevant traits such as altruism. DeGrazia also discusses an entirely speculative possibility which would entail developing an "artificial chromosome that includes multiple genes coding for stronger dispositions to a variety of moral virtues" (2014:362).

Walker, another supporter of moral bioenhancement, does not delve too far below the surface regarding scientific considerations. He posits that for moral bioenhancement to be achievable, it would require interdisciplinary involvement. Psychologists, could identify 'virtuous' personality traits, whereas behavioural geneticists could track the extent to which such traits are genetically influenced (Walker, 2009:31). If such traits are identified, there would either be the option of pre-implantation selection of embryos displaying the requisite moral traits or genetic interventions could be performed to 'switch off' genes associated with negative moral traits. Walker also considers the possibility of inserting artificial genes, mentioned by DeGrazia. In terms of the foundations that would have to be established for a project of moral bioenhancement to be feasible,

[20] The genetic basis of counter-moral traits is based upon identical twin studies research. Twin studies indicate that "antisocial and aggressive behaviour has a considerable genetic basis…[the level attributable to genetics is placed] at between 40-60%" (Glen & Raine, 2013:54). However, it is most likely that a number of "gene variants" are involved with a genetic predisposition to aggressive behaviour rather than a single gene (ibid.). Further support for the biological basis of morally relevant dispositions or emotions is provided by Van Goozen and Fairchild who have also explored "the biological correlates of antisocial behaviour in children" (2008:942).

Walker outlines the three propositions that would have to be empirically verified. Firstly, it would have to be shown that there are in fact definitive "character traits…[that include] virtues (or vices)" (Walker, 2009:31). Secondly, it would have to be ascertained that "at least some virtues (or vices) have a heritable component" (ibid.) and thirdly, that we are able to then "detect and control the genes responsible for this heritable component" (ibid.).

Walker argues that the first and third propositions would not be easily settled as this would require resolving disagreements between situations and personality theorists (2009:31). Situationists are of the view that behaviour is wholly, or at least substantially, influenced or determined by environmental factors; whereas personality theorists argue that personality, a product of genetic processes, is the foundation of behaviour (Walker, 2009:31-32). In common parlance, this debate is often referred to as the *nature vs nurture debate*. For moral bioenhancement to be feasible, Walker argues that personality theorists would have to conclusively establish their position as correct (2009:32). In this regard, Walker discusses research, including the identification of a gene associated with "novelty-seeking" (ibid.) that provides tentative grounds for support of the personality theorist view (Ebstein et al., 1995; Benjamin et al., 1996). Based on animal studies (Menzel, 1974; De Waal, 1996; Hamilton, 1964; Trivers, 1971), virtues with genetic components that Walker examines as a potential basis for a project of moral bioenhancement are: *truthfulness*, *justice* and *caring* (2009:32-34). However, Walker, is careful to avoid a charge of genetic determinism and points out that "genes *influence* but do not *determine* personality" (2009:38). Therefore, possessing a gene associated with any one of the above virtues will not ensure that this virtue is manifested in the individual's behaviour.

Spence is a more indirect supporter of the aims of moral bioenhancement in that he published an article that predates the first 2008 Persson and Savulescu and Douglas articles that discuss moral bioenhancement. Spence posits that, in a certain sense, the field of psychiatry is already engaged with a form of pharmacological, and thus biological, "moral assistance" (2008:179). In other words, the treatment of pathological psychological functioning generally results in behavioural improvements that have moral implications. Akin to the point made by Harris and mentioned in footnote 19, the prima facie response to this could be to dismiss such interventions as aimed at bringing about behaviour that has moral consequences rather than moral behaviour itself, or, to dismiss such interventions as artificial in nature, and thus, as not constituting true moral enhancement (Spence, 2008:179). This would be congruent with the line of argumentation that regards morality as inhering in the means taken and the effort made to achieve moral ends.

However, Spence is of the view that whether or not an intervention can be classified as a moral enhancement "depends crucially upon the goals of the patient concerned, i.e. what are the 'ends' that he is pursuing?" (2008:179). Patients with pathological psychological conditions, often take medication to improve their behaviour and conduct not only to assist in their own well-being but also to lessen the negative impact of their behaviour on loved ones. Generally, if an individual is motivated by such a goal, and in the process, acts in such a way as to ameliorate or improve his behaviour, we would unequivocally view this as a form of altruism and thus as moral behaviour (Spence, 2008:180).

Like Spence, Pacholczyk is not an explicit supporter of moral bioenhancement as construed by its major proponents. However, she argues that the negativity regarding the possibility of moral bioenhancement can be attributed to the fact that expectations for what it could achieve are unrealistically high (Pacholczyk, 2011:160). Pacholczyk does not think moral bioenhancement will be able to avert ultimate harm or address the kinds of problems that Persson & Savulescu expect of it (2011:168). She points out that we currently treat a number of minor and more serious psychological conditions with pharmacological interventions. Furthermore, we have realistic expectations of the efficacy of such treatments. If we view the potential efficacy of moral bioenhancement in a more realistic manner, akin to how we view pharmacological interventions, then moral enhancement may be possible (Pacholczyk, 2011:170). Pacholczyk discusses the use of oxytocin as a mechanism of moral bioenhancement that warrants further research (2011:173).

As mentioned in section 2.2, Douglas associates moral bioenhancement with the reduction of interference in good motives and argues that despite moral pluralism there are certain "counter-moral emotions" (2008:231) that would be unanimously viewed in a negative light. In this regard, he identifies "a strong aversion to other racial groups…[and] the impulse towards violent aggression" (2008:231), as the two primary counter-moral emotions that could be candidates for moral bioenhancement. Whilst he agrees with the view that our moral psychology is of such a complex nature that it may never be possible to biologically modify certain aspects of it, he doesn't believe that this is necessarily the case with two above-mentioned counter-moral emotions (Douglas, 2008:233). Research in the fields of neuroscience and behavioural genetics indicate that aggression has a genetic basis (Crowe, 1974; Cadoret, 1978; Grove et al., 1990; Brunner et al., 1993; Caspi & McClay, 2002; De Almeida et al., 2005) and fMRI scans have shown that certain brain activities play a role in racial aversion (Hart et al., 2000; Phelps et al., 2000; Cunningham et

al., 2004). He argues that given the fact that we have tentatively identified these biological associations, this should give us hope that future progress will be made in this regard.

2.4.2 The response of the sceptics

As mentioned above, there are a number of thinkers who address the scientific feasibility of moral bioenhancement. De Melo-Martin and Salles provide one of the strongest claims against the scientific plausibility of moral bioenhancement. They argue that the supposed scientific evidence discussed by supporters of moral bioenhancement is actually "highly contested…[and] grounded on problematic presuppositions" (De Melo-Martin & Salles, 2015:229). They see most of the supposed scientific evidence as having been affected in this way; however, they only discuss the phenomenon of racial bias which is alleged by supporters of moral bioenhancement to be biologically mediated. Douglas, in particular, has cited studies indicating that the amygdala is activated when feeling the 'emotion' of racial aversion. However, De Melo-Martin and Salles argue that in order to posit that it is possible to attenuate this counter-moral emotion, one has to make a number of presuppositions which amount to question-begging. One has to assume "that racial aversion is indeed an identifiable emotion that can be manipulated, that the amygdala's activation constitutes a response to an aversive emotion, and that the amygdala regulates in particular the affective component of the emotions" (De Melo-Martin & Salles, 2015:230). While the onus is on Douglas to present an argument for the above three assumptions, De Melo-Martin & Salles see him as simply assuming their validity in order to lend support to the scientific feasibility of moral bioenhancement.

A second group of thinkers provide arguments that can be loosely categorised as stating that the scientific claims made in the moral bioenhancement debate are, at this point, overstated. In other words, thinkers advancing such arguments are of the view that given our current level of scientific ability, as well as certain other constraints, it is more likely that moral bioenhancement will never be a safe or effective possibility. One of the most frequently cited proponents of this view is the neuroscientist Molly Crockett. To be fair, however, Crockett's view regarding the scientific feasibility of moral bioenhancement may be viewed as more neutral than her inclusion in this section would imply. Crockett firstly points out that research in the field of genomics does not provide support for the likelihood that particular moral dispositions, such as altruism for example, have single gene origins (2014:370). She then discusses how her work regarding the effects of SSRIs as a means of countering aggression has been cited by a number of supporters of moral bioenhancement such as DeGrazia (2014:361). Crockett points out, however, that research into

moral bioenhancement is extremely new and that her results are more complex than DeGrazia's use of them seems to imply (2014:370). Her research indicates that the SSRI *Citalopram* seems to decrease an individual's tendency to cause harm; however, this research was conducted using hypothetical situations, it has not yet been replicated "in the laboratory in healthy volunteers" (Crockett, 2014:370). Therefore, whilst her findings provide tentative 'proof' regarding the ability of SSRIs to mitigate aggression, more research is needed.

Furthermore, even if the effects of SSRIs on aggressive behaviour could be established, it would be necessary to fully understand their effects on other physiological functions. Crockett argues that serotonin affects not only "social behaviour…[but also] plays a role in a variety of other processes, including (but not limited to) learning, emotion, vision, sexual behaviour, appetite, sleep, pain and memory, and there are at least 17 different types of serotonin receptors that produce distinct effects on neurotransmission" (2014:370). Jones has also discussed the fact that serotonin influences a number of other areas such as "cardiovascular regulation, respiration, sleep-wake cycles, and reward learning" (2013:3). Therefore, there is concern that adjustments aiming to target moral behaviour may negatively affect these areas (ibid.). This indicates that risks and side-effects must be afforded considerable attention when considering the possibility of moral bioenhancement via the use of SSRIs. Of course, as Crockett points out, scientific progress will presumably address such side-effects (2014:370). In terms of the different targets of moral bioenhancement that have been proposed in the literature, Crockett argues that "from a neuroscientific perspective, the evidence so far suggests that targeting moral motivation may the most promising avenue for promoting moral behaviour" (2014:370). This view is supported by research conducted on psychopaths indicating that their pathology does not lie in their inability to recognise that certain actions are morally problematic; but rather in their possessing inadequate moral motivation (ibid.).

Of course, it could be pointed out that we already make extensive use of SSRIs to treat conditions ranging from mild to severe depression, and, that we are therefore already engaged in a form of treatment that has moral implications. Sparrow however disagrees with the claim made by supporters of moral bioenhancement that softer forms of moral bioenhancement, such as "drug therapies", are already available. He doesn't see this as a form of moral enhancement as it is not self-evident to him that the effects of drugs can be equated with being "more moral" (2014:20). In a similar manner to Agar and Harris, Sparrow argues that being more moral requires "means-end reasoning" (2014:22). Here, he is, of course, alluding to the point that has already been discussed,

namely, the view that morality is akin to a verb rather than a noun. In other words, it inheres in a process rather than in a goal, and reasoning is integral to this process. In addition, Sparrow sees some of the arguments in the moral bioenhancement debate as guilty of utilising *shonky* science that is based on outdated and controversial sociobiological claims (2014:27). The overall assumption that individual morality "is a function of [a] person's neurochemistry and/or that person's genetics" is for him problematic (Sparrow, 2014:27). Sparrow argues that this assumption supports the interpretation that those "who are immoral, are incorrigibly so, while those who are most moral are good by nature" (2014:27)[21].

There are a number of thinkers, including scientists, who argue that moral bioenhancement will, most likely, not be possible – or would be inadvisable, even if possible – due to the complexity of human moral psychology. Young and Duncan argue that in the past it was posited that moral cognition is *domain-specific*, referring to the view that specific areas of cognition and particular neural processes could be pinpointed as responsible for moral decision making (2012:1). However, the view is now that moral cognition is *domain-general*. In other words, moral decision-making is the product of complex processes diffused throughout the brain. Therefore, if we ask the question: "where in the brain is morality" (Young & Duncan, 2012:1); the answer is both "everywhere…[and] nowhere" (ibid.:7).

Zarpentine argues that for moral bioenhancement to effectively achieve its aims, it must be able to target specific areas without negative side-effects (2013:145). However, due to the "ontogenetic and neuropsychological complexity exhibited by human moral psychology" (2013:145), Zarpentine doesn't see this as possible in the immediate future, although he does not state that it may not be possible in the future (ibid.:150)[22]. Lecher argues that, in their claim that "a neurochemical substance can bring about a motive in the human mind, and that this motive suffices to push the individual into action" (2014:31), supporters of moral bioenhancement are providing a reductionist argument. Consciousness leads to action along a chain which includes:

21 It must, however, be noted that Sparrow is rather unfair here as supporters of moral bioenhancement, such as Persson and Savulescu, have explicitly stated that they are of the view that morality is in part or is at least "significantly…[rather than solely] a function" of our biology (Persson & Savulescu, 2014b:42). They frequently state that moral bioenhancement should be complemented by traditional forms of moral enhancement or education (Persson & Savulescu, 2012:11).

22 Zarpentine interprets ontogenetic complexity as indicative of "complexity in moral psychological development…[referring specifically to] complex interactions between genetic inheritance and environment in moral development" (2013:145).

> brain states, mental states (thoughts), intentions (beliefs, desires, fears), propositions (statements), actions [and] settled dispositions to act (character traits). The farther an element is located from the chain's origin, the more complex the account of this element becomes. Exponents of moral bioenhancement, [like Douglas and Persson and Savulescu] want to tell a story about the middle part of the chain, about (moral) motivation, and sometimes about the end of the chain, about (moral dispositions), but their explanans are drawn from the origin (from the category of brain states), leading them into the trap of reductionism (Zarpentine, 2014:31-32)

Bronstein argues that even if it could be definitively shown that moral dispositions or virtues are biologically influenced, specific moral traits are almost certainly not informed by single genes; rather, what is far more likely is that such traits are polygenetic (2010:85). In other words – this is the point also mentioned by Crockett – genes associated with specific moral traits may also be associated with other physiological, non-moral processes and such moral traits may, in turn, be associated with multiple genes. Therefore, altering such "polygenes may threaten many systems within an organism" (Bronstein, 2010:85) and result in devastating side-effects. Sprinkle (2010), Andreadis (2010) and Arnhart (2010) critique Walker's argument in support of moral bioenhancement on similar grounds. Andreadis argues that there is no evidence shwing that the kind of moral behaviour discussed in the literature is the product of single genes (2010:76). Rather, the prevailing view is to recognise that genes produce "*pleiotropic* (i.e. multiple) effects" (ibid.). In order to safely alter such genes, one would need extensive knowledge regarding the way in which they interact, which we do not currently, or even approximately, possess (Andreadis, 2010:76).

While Andreadis argues that, within the field of biology, the important influencing role of genes on behaviour is now accepted as definitive; the view is that we have moved beyond the former polarity of the nature/nurture debate. The prevailing view is now a "gene/environment (G x E) interplay model" (Andreadis, 2010:76) which makes the task of linking specific moral dispositions to genes an outdated one. Furthermore, regarding the implications for moral bioenhancement, due to this above-mentioned complexity, the view is that "it is far more feasible to correct an error than to 'enhance' an already functioning brain" (ibid.). Arnhart also mentions the complexity of environmental and genetic interaction as well as the view that the risks of intervening in genetic processes are such that it may never be safe, or worthwhile, to do so (2010:80). Sprinkle supports a similar conclusion by arguing that given problems related to the polygenetic influences on moral dispositions, moral bioenhancement will never be feasible; and on the off chance it would be, it wouldn't be morally justifiable (2010:88).

Other thinkers focus more on arguing against moral bioenhancement due to the conviction that side effects, risks or negative consequences in general will be too great. While there is much focus on the possibly negative side effects of genetic intervention, Clausen discusses the identity altering nature of deep brain stimulation (DBS) which has been mentioned as one of the possible mechanisms of moral bioenhancement. DBS is currently utilised for therapeutic interventions. However, due to the possibility that DBS may result in personality changes, as well as other risks, Clausen sees it as an option that is ethically unjustifiable, for enhancement purposes, at present (2010:1159). This concern will be discussed in detail in chapter 4a.

Neuroscience has identified certain morally relevant traits which "include sympathetic feelings, a primitive sense of right and wrong, a general sense of rules, highly self-conscious shame reactions, and effective self-control (ie power of will)" (Barilan, 2015:79). However, Barilan argues that these morally relevant traits must be flexible enough to ensure that they are able to be balanced by "culture and self-reflection [which] integrate all relevant factors and produce a morally desired behaviour" (2015:79). An emphasis on the enhancement of particular traits may "come at the expense of complementary ones…[as well as] at the expense of a psychological flexibility that is a moral sub-capacity in its own right" (Barilan, 2015:83).

Connected with a focus on the complexity of human moral psychology and the concern for risks and side-effects, are arguments that posit that moral bioenhancement will morally worsen us. Agar argues that moral bioenhancement is dangerous for similar reasons to Barilan (2015a). He argues that our moral judgement is founded upon "cognitive, emotional and motivational sub-capacities" (Agar, 2015a:343) and that isolating and strengthening just one area or sub-capacity will result in imbalances. In other words, "with respect to moral sub-capacities, excesses are as bad as deficiencies. The unbalanced strengthening of a moral sub-capacity can lead to the endorsement of moral ideas that we rightly reject" (Agar, 2015a:344). In particular, Agar argues that strengthening certain emotions, such as empathy, could make it more likely that our reasoning capacities are bypassed, and this, he speculates, will result in a stronger tendency to make morally problematic consequentialist judgements (2015a:345). Drake has responded to Agar's argument in this regard. He points out that because Persson and Savulescu have explicitly stated that they are only aiming to increase moral motivation to the highest existing levels within the population – i.e. they are not aiming at radical enhancement – Agar's fears are exaggerated (Persson & Savulescu, 2012:113; Drake, 2016:5). Drake points out that if we look at individuals who are considered to be morally exemplary, such as Archbishop Desmond Tutu, who is described as

having extremely high levels of empathy, we can see this has most certainly not resulted in him having "dangerous moral judgement" (ibid.); in fact, the opposite seems to be the case.

Harris has also argued that moral bioenhancement, as construed by its supporters, could negatively impact our morality. I refer here not only to his concern for the subversion of moral autonomy, which will be discussed in detail in chapter 4b, but more specifically, to his discussion of the connection between serotonin and moral behaviour. Similarly to Agar, Chan and Harris argue that contrary to what is argued by supporters of moral bioenhancement, serotonin could play a negative role in moral behaviour in that they see it as heightening emotional, rather than rational, responses to morally relevant situations (2011:130)[23]. Chan and Harris discuss how serotonin purportedly increases an aversion to violence, thus reducing the likelihood of an aggressive response in any given situation. However, they argue that in certain situations the morally right course of action may require an aggressive response. Here they mention Jasper Schuringa who used violence to overpower an attempted plane-hijacking, thereby saving the lives of all on board. Chan and Harris argue that if Schuringa had been morally bioenhanced against aggression he may not have attempted such a heroic feat (2011:131).

Another example that has been given of a context in which proposed enhancements would be at odds with the requirements of specific situations is that of a judge who, it is argued, would not be able to fulfil her job appropriately if her empathy levels were boosted. Wasserman argues that for those "in positions of power and authority" (2014:374) in general, an enhanced sense of empathy may be at odds with the requirement to act in a way that, while morally problematic, is for the greater good, all things considered. Casal also discusses this issue and points out that viewing moral bioenhancement as a means of boosting the motivation of individuals to do what they know is right, could backfire momentously (2015:340). For example, a terrorist considering a suicide bombing may acquire the 'courage' to undertake such an act if his sense of altruism and empathy for those in whose name he undertakes such an act is intensified. Casal argues that it is dangerous

[23] Of course, as Bublitz points out, it is correct that serotonin plays a role in emotional regulation, as evidenced by its association with various disorders of affect (2016:91). This is, however, inferred on the basis of indirect evidence, namely, through neurobiological changes in areas of the brain that are associated with emotion. Nevertheless, it is likely that using serotonin as a means of moral bioenhancement would work in two possible ways. It could intensify certain "emotion[s] at the expense of another" emotion (ibid), such as intensifying empathy for another individual at the expense of emotions associated with a regard for the self. In this way, there could be minimal impact to cognition and deliberative processes. On the other hand, it could intensify certain emotions with the effect that the strength of such increases could overwhelm cognition and deliberation. In this way, as Bublitz points out, "the reasons in favour of action A…[would] no longer outweigh the emotions striving for action B" (2016:91-92). It is the second outcome that Harris fears will be produced by moral bioenhancement.

to assume that enhancing our levels of altruism and sense of justice "in the absence of just institutions" (2015:340) will not bring potentially great risks.

However, Kahane and Savulescu have responded to Harris's argument by arguing that his argument against serotonin as a moral enhancer is not an argument against moral bioenhancement; "it is an argument against raising serotonin" (2015:135). If it could be shown that the administration of serotonin did not produce the requisite moral effects or behaviour, then, of course, other methods of moral bioenhancement would be preferable. However, this remains to be seen. The issue that is problematic in the Chan Harris line of argumentation is their claim that increases in empathy, and thereby emotions, will automatically bypass reasoning or lead or morally problematic consequentialist-type judgements. In this regard, ground-breaking scientific research utilising functional magnetic resonance imagining (fMRI) to investigate the neurobiological influences of moral judgements has indicated evidence to the contrary (Greene, et al., 2001).

This research entails using fMRI to analyse what occurs in the brain when individuals are presented with moral dilemmas, such as trolley problems. In trolley problems, an individual must imagine a train headed towards five individuals who are tied to the track with no means of escape. The only way to save these individuals is to pull a lever, thereby diverting the train to an alternate track upon which a single individual is restrained. When faced with the decision of whether or not to pull the lever, most individuals answer that they would pull the lever as it is preferable to save more lives where possible, even if one life is 'sacrificed' to do so. This is a decidedly utilitarian judgement as it judges the correct course of action to be that which produces the best consequences in terms of the maximisation of overall utility or social well-being. The second part of the trolley problem entails imagining a single train track with five individuals, once again, tied to the track. In this version, one is standing on a bridge overlooking the track and has the option of pushing a large stranger, who is also standing on the bridge, onto the track, thereby halting the progress of the train and saving the lives of the five. Generally, most individuals react with great discomfort when faced with this second dilemma and opt to not push the stranger onto the track, despite the identical rationale of saving more lives being present.

Responses to the second dilemma are associated with deontological judgements; namely, the wrongness of using an individual as a direct means to achieve a particular end, even if this end is positive in that more lives will be saved. Greene et al argue that the difference between the two dilemmas is that the second, deontological judgement, elicits a strong emotional response whereas

the first, utilitarian, and thus consequentialist, judgement, is a more calculative or rational response (2001:2106). Their hypothesis was affirmed by experiments indicating that brain areas associated with emotion were far more active when making deontological judgements than consequentialist or utilitarian ones. However, as discussed in the previous section, this dispute between supporters of moral bioenhancement, such as Persson and Savulescu et al., and their opponents, is indicative of the deeper meta-ethical dispute regarding the extent to which morality should be a product of emotions or rationality; and which moral theory better serves these ends: consequentialism or theories that emphasise means/end considerations. This matter is of great relevance to the question of whether moral bioenhancement will subvert moral autonomy and will be therefore be addressed further in chapter 4.

2.5 Concluding remarks

In this chapter, I presented three interconnected areas that I identified from a survey of the literature, and take to be integral to the coherence of the moral bioenhancement project. I framed the three areas as problems due to the fact that they are characterised by disagreements that would have to be resolved for moral bioenhancement to become a coherent possibility. In terms of the first problem, various definitional disagreements regarding moral bioenhancement are related to the underlying challenges and disagreements pertaining to the second problem, namely, the identification of the content of morality itself. In other words, to define moral bioenhancement one must be able to stipulate what the target/s would be, and, this latter stipulation is dependent upon one's conception of what is salient regarding morality itself. In addition, the problem of science also requires clarity and consensus regarding the target/s of enhancement for moral bioenhancement to be a scientifically viable possibility.

The problem of the science of moral bioenhancement also introduces its own novel concern; namely, the fact that whatever target/s of morality we have identified must be susceptible to safe biological modification. The debate here has been dominated by Persson and Savulescu's identification of altruism and a sense of justice as two morally relevant dispositions that are potentially susceptible to biomedical interventions. However, there are a variety of arguments provided by those who are decidedly sceptical regarding the scientific feasibility and risks posed by moral bioenhancement due to the complexity of human moral psychology, and thus, the possible side-effects. The various arguments in this section do not provide conclusive 'evidence' for or against the claims that moral bioenhancement will be a scientific possibility, as such evidence is impossible to ascertain, due to the fact that the science behind moral bioenhancement is

predominantly a speculative matter. However, in this dissertation, I will operate with the idealising assumption that moral bioenhancement could be safe and effective at some point, in order to investigate the deeper ethical and philosophical concerns that it elicits

In terms of the related problem of moral content, much attention is focused upon what the broader target of moral bioenhancement should be: affective or cognitive capacities. I would argue, however, that this dichotomy is a false one as morality is far more complex in character than this polarisation would imply; both do, and should, play a vital role in moral behaviour. This view is supported by neuroscientific research, as discussed at the end of section 2.4.2. Thus, I would argue that any attempts to define morality or moral bioenhancement, must include some reference to both affective and cognitive components, even if this is implicit. In addition, I would argue that the Socratic posit that knowing the good implies doing the good represents an idealised account of human nature which is at odds with an abundance of empirical evidence to the contrary. I do not regard the primary moral problem of the twenty-first century to be confusion regarding what is moral and what is not. Rather, I concur with Persson and Savulescu that the primary weakness of existent human morality is a lack of motivation, or insufficient desire, to carry what is known to be good into action. I would argue that this 'problem' is deeply connected with a general lack of willingness to consider the perspective of others as a morally relevant factor in determining one's behaviour, particularly where doing so would require any form of sacrifice. Thus, I take motivation to be an appropriate target for moral bioenhancement.

To conclude this chapter, I will now provide a working definition of moral bioenhancement that will inform my investigation in this dissertation. Whilst, as observed in the literature, definitions of moral bioenhancement do tend to be more normative than descriptive, I would argue that this is unavoidable. Morality is, by definition, a normative matter, and in this regard, moral bioenhancement, which signifies the attempt to improve morality, is also a thoroughly normative endeavour. In terms of nomenclature, I will purposely utilise the term *moral bioenhancement* in this dissertation, rather than neutral terms such as biomoral modulation, modification, alteration, manipulation or change, in order to emphasise that I am referring to moral improvement, by definition. This signifies that an intervention could only be referred to as a moral bioenhancement if it resulted in a discernible, and stipulated, improvement in conduct or behaviour.

The definition that I will offer is, therefore, one that I posit would constitute a definitive moral improvement, if it were to be realised. Whether or not the aims described in the definition are

scientifically possible is, of course, another matter. I would also argue that a definition of moral bioenhancement must take a position regarding what should be the target of interventions. Different targets that have been discussed in the literature are: (i) moral behaviour or action; ii) motives, will or intentions; (iii) moral dispositions, capacities or attitudes, such as a capacity for altruism or empathy and a sense of justice, as well as those counter-moral emotions that may interfere with our motivation to act in a moral manner; and, (iv) moral decision-making, judgement or reasoning I regard all these targets as relevant to morality, and thus, to a definition of moral bioenhancement.

The use of the term bioenhancement implicitly distinguishes biological moral enhancement from traditional moral enhancement, and is thus, a narrow or more specific definition. I agree with arguments claiming that the long-term influence exerted by both biological and traditional moral enhancement could be comparable, however, while some thinkers claim that there is no morally relevant difference between the two as they are both aiming at 'the good', I would posit that they are different in kind, due to the means that they employ to achieve their ends. Here, the distinction between active and passive enhancements is of great relevance and will therefore be discussed further in chapter 4. In addition, whilst I would opt for a definition that avoids explicit reference to the treatment/enhancement distinction, due to various problems associated with it, I would argue that there is merit in including a reference to some maximum level of moral functioning that would be desirable. In other words, if the aim of moral bioenhancement is simply to raise the level of moral functioning to the height displayed by those members of society regarded as moral exemplars, this would allay fears regarding radical enhancement. In this regard, a definition should make it clear that the aim would not be to boost moral functioning to unspecified levels, but rather, to elevate it to a level that is already present in the human population.

With all of the above in mind, I define moral bioenhancement as:

> A biomedical intervention that results in a discernible improvement in considered moral conduct to levels displayed by moral exemplars within society, where this improvement is achieved by way of increasing the motivation of individuals to act in a morally desirable manner, by either raising low levels of morally relevant dispositions or mitigating high levels of counter-moral dispositions. The primary target would be the disposition of reflective empathy, where this would entail improving an individual's ability to consider the perspective of others as a morally relevant factor in determining their actions.

This will be the definition that informs my discussion and investigation in this dissertation. It is clearly informed by aspects of both Persson and Savulescu and Douglas' interpretation of moral bioenhancement, however it differs regarding the stipulated levels of improvement, as well as the

notion that moral conduct must be *considered* and empathy must be *reflective*. This implies that improvements must be accompanied by the ability to provide *authentic* reasons for changes in behaviour. I have included empathy in my definition, as I posit that it would be a suitable disposition to focus on as a target for enhancement. This is firstly, because there is evidence that it is susceptible to biomedical mediation. However, whether this will ever be scientifically possible, in the way that is hoped is, of course, far from certain. Secondly, as Persson and Savulescu have pointed out, empathy, or altruism, is a disposition that has a place in most moral theories (2012:403). Where empathy is associated with the ability to consider the perspective of others as a morally relevant factor in determining action, this disposition is closely associated with a sense of altruism or selflessness. Utilitarianism requires one to set aside one's own interests in order to consider the wellbeing of others, which is why it is often regarded as an overly demanding moral theory. In addition, acting from duty, and in accordance with the moral law, as espoused by deontology, would also require that one disregards one's personal preferences or feelings, in favour of doing what is right, simply because it is the right thing to do. Thus, in both utilitarian and deontological moral theories – generally viewed as polar-opposites – it can be argued that a sense of selflessness is required. However, the difference between the two theories lies in the justification for why this should be the case. I will utilise this proposed understanding of moral bioenhancement in subsequent chapters.

Chapter 3 – In-practice objections to moral bioenhancement

3.1 Introduction and overview of chapter

Arguments that are given in support of moral bioenhancement are generally consequentialist in nature. In other words, they provide reasons for support of moral bioenhancement as a means to some beneficial end or the avoidance of some harm. As outlined in chapter 1, examples of harm avoidance would be the implementation of moral bioenhancement to mitigate against global catastrophe, termed ultimate harm by Persson and Savulescu, or to simply make us 'better', and thus more willing, and able, to address various global problems that are viewed as inextricably connected to deficiencies in human moral psychology. As discussed in the previous chapter, most supporters of moral bioenhancement make a fundamental distinction between traditional – non-biological – moral enhancement and biological moral enhancement and argue that the former is inadequate, alone, as a means of addressing the variety of urgent problems facing humanity. A stronger argument could also be made that the problems and risks faced by twenty first century humanity are such that we have a moral duty, or obligation, to make ourselves morally better through moral bioenhancement. The main argument provided in support of moral bioenhancement by Persson and Savulescu was presented in chapter 1 and will therefore not be discussed in detail in this chapter which will primarily address arguments that have been lodged against moral bioenhancement – and the responses to these arguments – in the literature.

The arguments against moral bioenhancement are, not surprisingly, rather extensive and varied. A useful way of categorising arguments, in this regard, is to distinguish between those that are *in-practice* objections and those that are *in-principle* objections[24]. In-practice objections admonish us to take cognisance of the practical negative impact that moral bioenhancement may have on, and within, the real world (Agar, 2015: 38). In other words, such objections do not focus on the moral status of moral bioenhancement itself, but argue that the project may fail to realise its intended, positive aims, and, in this way, will leave us worse off. In-practice objections are thus consequentialist in nature. Whilst, in the context of moral bioenhancement, such concerns are

[24] This is a helpful distinction suggested by Agar (2015:38) that avoids some of the problems associated with distinguishing between instrumental versus intrinsic concerns against moral bioenhancement. Instrumental concerns are related to in-practice objections, in terms of their shared focus on the possible negative effects of moral bioenhancement, whereas arguments that focus on the intrinsic wrongness of moral bioenhancement would claim that the practice is wrong, in and of itself, for various reasons, regardless of any benefits it may produce. While this distinction may also be helpful, it disregards the areas and concerns where there may be overlap. For example, the concern that moral bioenhancement will impact negatively upon morality is both an instrumental concern, due to the potentially negative consequences this may produce, and a concern for the intrinsic wrongness of impacting upon something that is viewed as having absolute value.

speculative in nature, they are nevertheless crucially important in terms of the potentially serious harms that they draw attention to. In-principle objections, on the other hand, make the claim that moral bioenhancement is wrong, in and of itself, – for some reason – regardless of any alleged good or avoidance of harms that it may produce. Many of these arguments draw upon insights from non-consequentialist theories.

Whilst I have grouped the arguments against moral bioenhancement into either in-practice or in-principle objections, these categories are, of course, not mutually exclusive. The different areas of focus are interconnected in that assumptions made in certain areas will impact upon how other areas are interpreted. In-principle objections have in-practice implications and vice versa. In addition, when looking at the claims against moral bioenhancement, it must be noted that the various arguments are predicated upon the way in which moral bioenhancement is defined; which, in turn, is influenced both by the moral theory that is subscribed to and other meta-ethical perspectives, such as what is viewed as constituting morality or moral content. These interpretations, in turn, will influence matters such as whether or not moral bioenhancement is regarded as feasible, and if so, what is targeted for moral bioenhancement. In this chapter, I will focus on explicating the most prevalent in-practice objections to moral bioenhancement. The primary in-principle objections to moral bioenhancement, which are the main focus of my dissertation, will be discussed in chapter 4.

In section 3.2 I will address some of the safety and risk concerns, for individuals, that moral bioenhancement poses, and, in particular, I will discuss concerns regarding possible negative side effects or harms that could result from 'tampering' with the biological foundations of our moral psychology. One such concern is that moral bioenhancement could negatively impact upon our morality, making it worse rather than improving it, and thus, could have the opposite effect of its intentions. In section 3.3 I will address other arguments that claim moral bioenhancement is unfeasible for various reasons. These concerns range from implementation and administration issues to logistical problems. Implementation issues generally discuss whether or not moral bioenhancement should be voluntary or non-voluntary, and thus, universally implemented, and the problems associated with both. This matter will be discussed in section 3.3.1. In section 3.3.2 I will discuss various administration problems such as: who will decide what should be enhanced, who will oversee this process and how the project could be protected from abuse, exploitation or nefarious intentions, such as the furthering of moral eugenics agendas. This concern can be termed *The problem of who will guard the guardians.* A related concern, discussed prevalently in the

literature, is *The bootstrapping problem.* This refers to the problem that it is morally deficient human beings, the targets of moral bioenhancement, who would oversee the process of moral bioenhancement and ensure that it occurs in an ethical manner (Persson & Savulescu, 2012:2-3). Other thinkers argue that moral bioenhancement implemented at the individual level will be ineffective in dealing with the potential catastrophes, such as the avoidance of ultimate harm, that are given as the reason for why we need it in the first place. Finally, in section 3.4 I will discuss distributive justice and egalitarian concerns related to who will have access to moral bioenhancement technologies and whether issues of access will exacerbate existing inequalities.

3.2 Concerns regarding potential harms, safety and risk to individuals

Concerns for potential harms regarding the safety, and thus, the risk, involved in moral bioenhancement interventions, are, of course, the most obvious point of contention between supporters and opponents of moral bioenhancement. These concerns focus on the possible negative side-effects that could result from 'tampering' with the biological foundations of our moral psychology. In other words, the concern is that in our attempts to morally improve ourselves, we will potentially achieve the opposite. This concern for the negative impact of moral bioenhancement on our morality has both instrumental and intrinsic components. I will deal relatively briefly here with the former as this area was addressed in section 2.4.2 of the previous chapter. The concern for the intrinsic wrongness of moral bioenhancement is predominantly reflected in arguments that address the issue of the impact that moral bioenhancement could have on human moral agency or autonomy, and thus, morality in general. This will be addressed in chapter 4

In their literature review, Specker et al. present areas of the moral bioenhancement debate that they regard as having been neglected (2015). They argue that whilst psychopharmacological interventions are frequently discussed as potential mechanisms of moral bioenhancement, "it is particularly worrisome that surprisingly little attention is given to side-effects, risks and safety-issues" (Specker et al., 2015:14), and, in particular, the risks associated with brain interventions. They emphasise that this neglect should be addressed and become a key issue in the moral bioenhancement debate. However, it must be pointed out that whilst they are the primary proponents of moral bioenhancement, Persson and Savulescu have explicitly included a caveat that it should only take place if its safety and efficacy has been established (2008:174). Furthermore, the lack of focus on such matters is not surprising given the fact that the investigation of moral bioenhancement has predominantly taken place in the philosophical and ethical realm, whereas the

matter of risk and potential side effects is a concern requiring scientific expertise. Nevertheless, a number of thinkers have expressed the concern that issues of safety and risk are glossed over by the proponents of moral bioenhancement.

Beauchamp has argued that Persson and Savulescu do not pay sufficient attention to problems of safety and risk possibly due to the fact that concrete reassurance in this regard would be speculative (2015:347). Persson and Savulescu discuss, at length, the "highly destructive, uncontrollable processes…and catastrophic risk" (Beauchamp, 2015:347) that could be unleashed in the wake of general scientific and technological progress as part of their argument necessitating moral bioenhancement. But, as pointed out by Beauchamp, this catastrophic risk could be taken to include moral bioenhancement itself. Beauchamp posits that Persson and Savulescu's downplaying of safety issues could also be attributed to the fact that they don't regard moral bioenhancement technologies as posing any novel risks over and above the risks that are shared with "all new forms of powerful innovation" (ibid.). Rather, they see the most pressing safety risks as lying in the problem of "how to secure wise and proper applications" (ibid.). Beauchamp disagrees with this position and points out that altering human genetic structure could create unanticipated effects that could only become apparent after the fact; and could achieve the opposite of what was intended. However, Persson and Savulescu respond to this concern by emphasising the cautious nature of their proposal and point out that their project of moral bioenhancement is a way of providing a solution to the problems they outline "which though risky in the beginning, gets more secure the longer we succeed in walking it" (2015:351).

As discussed in section 2.4.2, a prevalently mentioned safety concern is that the biological basis of our moral psychology is sufficiently complex that safe and successful moral bioenhancement would be highly unlikely or impossible (Young & Duncan 2012; Zarpentine, 2013; Lechner, 2014; Bronstein, 2010; Andreadis, 2010; Arnhart, 2010; Sprinkle 2010; Barilan, 2015; Agar, 2010; Blackford, 2010). In response to such concerns, Casal has argued that due to our lack of information regarding the "effectiveness, distribution or risks [of moral bioenhancement] – we cannot decide on its desirability…All we can currently say is that some biotherapies may be permissible and worth discussing" (2015:342). Jebari acknowledges the risks involved with attempting to alter "our evolved psychology" (2014:254) but nevertheless posits that any risks should be assessed by weighing up the advantages that could be secured – and presumably the harms avoided – through doing so. The Nuffield report holds a similar position that "the risks and

benefits of each particular means of enhancement must be assessed on the basis of empirical evidence where possible" (2013:165).

Several thinkers have given substantive content to the above concern regarding the safety of intervening in our complex moral psychology. The prevalent view here is that moral bioenhancement could negatively impact upon our morality, making it worse rather than improving it; and thus, could have the opposite effect of its intentions. Harris and Agar are the most noted proponents of this view. At this point in time, the debate has largely focused on the potentially negative effects of pharmacological interventions, and in particular, on the effects they have on neurotransmitters, as these are the most rudimentary forms of moral bioenhancement currently available. As discussed in section 2.4.2, Agar argues that "moral bioenhancement is considerably more dangerous than Persson and Savulescu suppose" (2015a:343). However, Agar does not claim that moral bioenhancement is wrong in-principle or that it is "intrinsically mistaken...[he argues that] biomedical means to this end violate no physical laws. Rather…[he] reject[s] the practical agenda of Persson and Savulescu" (Agar, 2015a:343). In other words, he thinks that in their attempts to improve us morally we will actually end up being morally worse-off. He argues that in the same way that fitting a more powerful and stronger bionic leg to an individual would have a major effect on her biomechanics in the absence of adaptations, so too will strengthening an isolated component of human moral psychology (ibid.).

Agar does not, however, oppose biological interventions in cases of pathological psychological functioning – moral therapy in other words – as he argues that in such cases, the moral judgement of such individuals is impaired due to a particular dysfunction, which, if corrected, would be far more likely to lead to improvement in overall functioning (2015a:343). To illustrate this point, he gives the example of how the moral judgement of a psychopath, affected by low levels of empathy, could be successfully treated and raised to levels of 'normalcy' (Agar, 2015a:343). However, in individuals who operate within the parameters of "moral normalcy" Agar is sceptical of the potential efficacy of moral bioenhancement due to the fact that "all of the cognitive, emotional and motivational sub-capacities that feed into moral thinking are working according to biological and psychological norms" (ibid.). It is the balancing of the above capacities that has produced what is viewed as "moral normalcy" which Agar points out, whilst not always correct, is nevertheless the "reference point for the [understanding], justification [and implementation] of ethical principles" (ibid.) regarded as morally defensible.

To illustrate this point, Agar discusses the bioenhancement of empathy. A claim made by Persson and Savulescu is that one of the deficits of human moral psychology is the existence of low levels of empathy for high levels of suffering experienced by large groups of individuals not in our immediate vicinity (2012:29). The supposed inability of those in developed nations to empathise with the suffering of the poor in developing nations would be an obvious example here. Persson and Savulescu argue that by administering oxytocin, "pro-social behaviour[s]" (2012:119), such as empathy, would be broadened and strengthened. Agar responds to this by pointing out that increases in empathy could occur either through intensifying the emotion experienced for one's immediate circle, one's in-group, or through extending it to include out-groups (2015a:344). There are, however, problems associated with both of these outcomes. Strengthening the experience of empathy for out-groups, or those removed from one's immediate vicinity, could occur at the expense of in-group empathy. In other words, it could result in what is often presented as one of the critiques of utilitarianism, namely, that it disrupts personal relationships, or, as the criticism is generally formulated, that it does not make provision for special obligations. Extending or strengthening our empathy, or pursuit of the good, for a general other could thwart an adequate observance of empathy for one's children or family, for example. In general, however, research has indicated that the administration of oxytocin produces the former result: a strengthening of existing feelings of empathy towards members of one's in-group.

Persson and Savulescu do acknowledge this and state that research has shown that "the pro-social effects of oxytocin are more accurately characterised as 'pro-in-group' effects" (2012:120). This is problematic because it is the enhancement of out-group pro-social effects that is posited as the solution to the problems that Persson and Savulescu present. They therefore admit that the administering of oxytocin to promote pro-social attitudes such as empathy would have to occur congruently with "reasoning which undercuts race, sex etc. as ground for moral differentiation" (Persson & Savulescu, 2012:120). However, they point out that the fact that a hormone such as oxytocin requires supplementation is not grounds for rejecting it altogether; it can still be extremely useful (ibid.). However, Agar argues that emphasising the role of reasoning in mitigating the strengthening of in-group empathy would not necessarily 'solve' the problem as our ability to arrive at morally desirable conclusions could still be at risk. Strengthening already existing empathy for our loved ones could lead to outcomes where we are no longer able to adequately weigh up competing options and come to sensible conclusions in the way that we currently do, by way of our existing moral capacities. It may lead to outcomes such as being more willing to impose suffering on out-groups, in order to protect, and thus favour, our loved ones from minor or trivial

suffering. In other words, the supplementation of oxytocin by reasoning that Persson and Savulescu discuss may achieve the opposite effect if the process of reasoning is influenced or overridden by the excessive strengthening of emotions such as empathy.

The example given by Agar to illustrate his argument is of a parent whose wish is to provide her child with the opportunity of an optimal education and considers breaking into a hospital and stealing a dialysis machine to sell on the internet to the highest bidder (2015a:344). Operating under the conditions of "moral normalcy", or the unhampered ability to weigh up competing harms and benefits, would lead the mother to realise that, as much as she loves her child and would wish to secure personal advantage for him, this course of action would be wrong due to its depriving countless individuals of the potentially life-saving benefits of the machine. Agar's point is that it is not evident that an individual with levels of empathy that have been boosted above normal levels would retain the ability to make such a judgement. Elsewhere, Agar articulates this concern slightly differently, arguing that moral bioenhancement could result in "a reduced sensitivity to moral reasons rejected by his or her enhancer" (2010:75).

Consequentialist theories, such as utilitarianism, place a high premium on happiness and the avoidance of suffering, whereas non-consequentialist theories, such as deontology, focus on the importance of the means taken to achieve a particular end, and the prioritisation of the role of acting in accordance with the dictates of rationality and the will. These different emphases can result in solutions to moral problems that are diametrically opposed to one another. Agar is of the view that an unenhanced utilitarian would at least consider deontological concerns in assessing a course of action; whereas it isn't clear that this would be the case if the bioenhancement of empathy had taken place (2010:75). In other words, Agar clearly envisages that a strengthening of empathy will lead to an over-heightened focus on reducing the suffering or increasing the happiness of loved ones, which will occur at the expense of our ability to employ moral judgement or cognition which enables us to weigh up competing harms and benefits and ensure means-end reasoning. In particular, his concern is that a strengthening of empathy could, in actual fact, result in a *distorted* utilitarian perspective in which an individual uses utilitarian reasoning but only applies it to his in-group rather than considering how his action could impact upon the utility of all individuals affected by it.

This is also the point that Harris makes when he interprets moral agency as requiring that one be able to make a judgement on what is the good or right action, "all things considered" (2016:28).

Furthermore, Harris argues that considering all things, "doesn't imply endless cost-benefit analysis, rather the assessment of the best course of action in most circumstances calls for the careful balancing of different sorts of harms, often to different groups of people" (2016:14). However, the obvious response to the above concern is to point out, once again, that Persson and Savulescu are not advocating *radical* enhancement of empathy. Rather, their argument is that elevating general levels of empathy in line with the levels displayed by the most empathetic members of society would constitute major moral improvement and alleviate some of the problems they outline. If the distribution of empathy in the population occurs in a bell curve shape then the aim would be to elevate the general population to the highest part of the bell curve; not above this. The point made by Drake, and mentioned in section 2.4.2, is also relevant here. If we look at members of society who display extremely high levels of empathy – the example he gives is Arch-Bishop Desmond Tutu – we would not argue that his ability to acutely empathise occurs at the expense of his moral reasoning (Drake, 2016:5). Rather, as Drake argues, his empathy "seems to facilitate exceptional moral judgement" (ibid.).

The concerns regarding the administration of serotonin to reduce aggression were also addressed in both section 2.4.1 and section 2.4.2, and will therefore not be repeated here in detail. The important point for the purposes of this section on side-effects and risks is to emphasise that while serotonin is currently extensively administered globally to treat cases of depression and anxiety, its effects on healthy individuals are not clear. As mentioned by Crockett, a neurotransmitter such as serotonin influences multiple brain processes, over and above its effects on reducing aggression. Therefore, an intervention that specifically targets a morally relevant behaviour such as aggression "by globally altering neurotransmitter function, may have undesirable side effects, and these should be considered when weighing the costs and benefits of the intervention" (Crockett, 2014:370). Furthermore, if a reduction in aggression may be equated with an increase in pro-social emotions, it is not evident, according to certain thinkers, that this is true moral enhancement. Rather, Harris argues that the use of a neurotransmitter such as serotonin is nothing more than "a policy of harm reduction by non-moral means by the administration of a molecule that reduces aggressiveness indiscriminately" (2016:83).

Harris, who equates 'moral goodness' with being able to make all things considered judgements, does not see the avoidance of harms, nor the elevation of pro-social behaviour, as constituent components of morality. Whilst reducing harmful behaviour through pharmacological interventions may give the appearance of moral improvement, "it is unlikely to change people's

moral outlook or judgements, although what happens certainly changes what they are able to do, or more modestly, what they are likely to do" (Harris, 2016:83). It is not only that pro-sociality is context specific and that pro-social behaviour exhibited towards other individuals may have negative effects for larger groups[25], but also, that Harris is of the view that pro-sociality is simply not "the stuff of which moral judgements are made" (2016:101). Harris argues that "there is reason to believe that increases in pro-sociality may not produce a morally better outcome 'all things considered' precisely because they reduce the ability of agents to consider all things" (2016:84).

In contradistinction to Harris' argument, however, Crockett et al cite their own research and evidence on the effects of serotonin to "promote pro-social behaviour by enhancing harm aversion, a prosocial sentiment that *directly affects both moral judgment* and moral behaviour" (own emphasis, 2010, Crockett et al.:2010a:17433). Elsewhere, in direct response to the challenges of Harris and Chan regarding their above claim, Crockett et al. charge Harris and Chan with a "narrow definition of moral judgment, one that is out of touch with empirical research" (2010:E184). They argue that Harris' account of moral judgement overemphasises the role of rational deliberation and fails to take into account the vital role played by "intuitive and emotional processes" (ibid.). Their claims are supported not only by experimental research[26] but also by the empirical observation that individuals frequently make moral judgements without being able to provide reasons to substantiate their conclusions.

3.3 Concerns regarding potential harms and risks to society

3.3.1 Implementation – compulsory or voluntary

The potential problems regarding the implementation and administration of moral bioenhancement address real-world practicalities, and thus, represent the most concrete sphere of the debate. The major issue in this area is the question of how a programme of moral bioenhancement could be

[25] What he means here is that in certain contexts, non-pro-social behaviours such as aggression or a lack of empathy are often the morally desirable and requisite behaviours. As mentioned in the previous chapter, Harris and Chan's example of the need for Jasper Schuringa to use aggression and force to overcome the hijackers of flight 253, thereby saving the lives of the 290 passengers on board, is one such example (2010:183). In this case, if Schuringa had refrained from aggressive engagement on an individual level, it would have resulted in decidedly negative effects for the individuals on board flight 253. Wasserman gives some other examples such as the requirement that certain professionals such as judges, and neurosurgeons, whilst operating, must neutralise their empathy, and, the need for ruthlessness, or lack of emotion, in certain highly pressured authoritative positions where decisions taken have major repercussions (2014:374).

[26] Research indicates that when individuals who have suffered from ventromedial prefrontal damage, with subsequent damage to "socio-emotional processes" (Crockett et al., 2010b:E184, Koenigs et al. 2007) are tested by being presented with moral dilemmas, their mechanisms of rationality remain functional whilst the afore-mentioned damage results in their displaying defective moral judgment.

implemented, thereby ensuring its success. The main questions here are: should moral bioenhancement be voluntary or non-voluntary, and thus universally implemented, and what are the problems associated with both?

Persson and Savulescu's argument for moral bioenhancement is predicated on the fact that the problems faced by twenty first century humanity are serious enough to warrant such a drastic solution. In other words, the avoidance of ultimate harm is deemed worthy enough to justify a programme of moral bioenhancement. In their earliest works on the subject, Persson and Savulescu argue that "if safe moral enhancements are ever developed, there are strong reasons to believe that their use should be obligatory, like education or fluoride in the water, since those who should take them are least likely to be inclined to use them. That is, safe, effective moral enhancement would be compulsory" (2008:174). Their conclusion here is inevitable, given their argument that humanity faces a grave risk of ultimate harm which could be perpetrated at the hands of a small group of individuals with nefarious intentions. If malevolent acts with destructive intentions can be exacted by such lone individuals or groups, with such catastrophic consequences, then clearly it will not do to make moral bioenhancement optional; such individuals would never volunteer to undergo such a procedure. Presumably those who would seek to undergo voluntary morally bioenhancement for themselves, or their children, would pose a minimal risk as perpetrators of ultimate harm. Thus, for moral bioenhancement to be an adequate solution to the risk of ultimate harm, it is an obvious conclusion that it must be mandatory, despite the unpalatable nature of such an enterprise.

Regarding the less immediate harms faced by humanity that are tied up with deficiencies in human moral psychology, there is also a strong argument in favour of compulsory moral bioenhancement. This is due to the fact that adequately addressing the global problems presented by Persson and Savulescu would require extensive, population-level 'correction' of these deficiencies. Furthermore, it seems unlikely that many would freely opt for moral bioenhancement without clear incentives to do so. This is a point alluded to by Douglas (2013:161) and Bronstein (2010:87). However, given the above discussion, whilst it may be evident how Persson and Savulescu come to the conclusion that moral bioenhancement would have to be mandatory to achieve its overall aims; their argument for compulsory moral bioenhancement represents the most obvious point of contention with critics and thus deflects adequate engagement with other important concerns. This could be the reason as to why, in a later publication, they retract the claim for compulsory moral bioenhancement. In their book on the subject, Persson and Savulescu argue that it "is [their] view

that some children should be subjected to moral bioenhancement, just as they are now subjected to traditional moral education" (2012:113). This is the sole allusion in the book to the issue of how moral bioenhancement could be implemented and appears to be a severe dilution of their earlier claim, and thus, seems to be a less efficacious solution to the problems faced by humanity that they have outlined.

Some have taken this omission to be evidence that Persson and Savulescu have changed their stance regarding the compulsory implementation of moral bioenhancement. In this regard, Rakić observes that "Persson and Savulescu diverge from their earlier position in no longer insisting on making moral enhancement compulsory" (2014:247). However, this omission clearly does not signify a change of heart on the part of Persson and Savulescu on the matter. When questioned regarding this seeming change in their position, Persson and Savulescu respond that they do not explicitly address the issue of whether or not moral bioenhancement should be compulsory, due to the fact that the technology that would enable moral bioenhancement is still in its earliest stages. Thus, they surmise that the issue does not require urgent attention until technological progress, in this regard, has been made (Persson & Savulescu, 2014:251).

However, in another article, published the following year, they seem to once again take a firm stance regarding the implementation of moral bioenhancement. Here, Persson and Savulescu explicitly argue that they "do not propose that moral bioenhancement should only be applied to those who voluntarily choose to undergo it, but that children should undergo it just as they have to undergo traditional moral education" (2015:350). They use an analogy regarding the current way in which we treat children for conditions such as attention deficit disorder by administering medications such as Ritalin in support of this proposal. They also provide examples of other conditions – such as the use of chemical castration for paedophiles – to argue for the permissibility of such a suggestion. This argument is weak, however, because the analogy between treating existing pathological conditions and the enhancement of species-typical levels of moral behaviour does not hold. It is also clear from other publications and comments made by Persson and Savulescu that their original view regarding compulsory implementation remains. In a later publication, Persson and Savulescu provide some support – at least in certain cases – for their earlier view. They argue that:

> *the risk* of humans causing ultimate harm if they do not undergo moral bioenhancement might not be so great that it is justified to make moral bioenhancement compulsory…However, in certain situations, the loss of freedom involved in making moral bioenhancement compulsory might be morally justified. Freedom is only one value and not the sole value; safety is another (Persson and Savulescu, 2014:251).

In a published debate with John Harris, when questioned on the matter of whether or not moral bioenhancement should be compulsory, Savulescu responds that whilst it would be far more straightforward to introduce the adoption of moral bioenhancement, or any new technology for that matter, as voluntary; "if the intervention is very effective and safe, and uncontroversially good, we should do it compulsorily" (2015:11). He adds that this would be akin to the general consensus that mandatory education for all children is correct due to the view that education "is clearly good for people and good for society" (ibid.). Later, in the same debate, it is pointed out to Savulescu that it seems that for the efficacy of his argument, moral bioenhancement would have to be implemented universally. Savulescu responds to this by arguing that this would be impossible, given the size of the world's population (2015:21). Rather, the most we could hope for would be to "change probabilities" (ibid.) regarding the risk of ultimate harm.

A number of thinkers have also addressed this point. Harris, in particular, has argued that to achieve the aims of prevention of ultimate harm, moral bioenhancement would have to "universal and exceptionless" (2016:137). However, it is most likely that in the same way that it is impossible to enforce the universal compliance of vaccinations, so too, would attempts at the universal administration of moral bioenhancement be impossible. Harris posits that this would be the case regardless of how easy and safe the process of moral bioenhancement could be made. Furthermore, the analogy is imperfect as vaccinations at least confer "herd immunity" (Harris, 2016:137), whereas the same would not be the case with the bioenhancement of morality. In this regard, Harris concludes that, regardless of the ethical problems associated with its aims, moral bioenhancement will not succeed (ibid.).

Trivino concurs that universal moral bioenhancement would be impossible due to technical constraints (2013:266). Furthermore, in the absence of universal adherence, we would be faced with situations akin to a prisoner's dilemma where unenhanced individuals could secure advantage at the expense of the enhanced (Trivino, 2013:267). This risk of exploitation would be a possibility not only between enhanced and non-enhanced individuals but also between enhanced and unenhanced groups or nations depending on the adoption of moral bioenhancement. Morioka concurs with the above thesis of the impossibility of universal implementation and discusses the possibility that affluent individuals would be more likely to have the means to avoid compulsory moral bioenhancement, and thus, compliance would be overrepresented by less affluent sectors of the population (2014:122). The potential for abuse and exploitation related to the non-universal

application of moral bioenhancement is a concern that will be discussed further below in section 3.4.

Of course, the most obvious concern with universal or compulsory moral bioenhancement is that enforcing it would amount to an infringement of individual freedom. In the moral bioenhancement debate there are two realms of potential threat to freedom. Firstly, *freedom of choice or action* may be lost if individuals are forced to comply with a programme of moral bioenhancement. Secondly, some thinkers argue that freedom, in the sense of *moral autonomy*, may be compromised or lost on a deeper level, even in cases of voluntary moral bioenhancement, due to how moral bioenhancement would work. This latter concern will be addressed in chapters 4 and 5. Regarding the concern for freedom of choice, one can respond in several ways. Firstly, one may provide a utilitarian response and argue that the benefits outweigh the harms, and that therefore, even if compulsory moral bioenhancement is enforced, with a resulting loss of freedom of choice, this compromise to freedom will be validated by the gains, namely, the improvements to human life in general, including the survival of the human species. This is the response of proponents of moral bioenhancement such as Persson and Savulescu. They firstly point out that human beings possess relative power rather than absolute power with regard to their freedom to perform any action imaginable. In this regard, "our power to act of our own free will is [already] a matter of degree" (Persson & Savulescu, 2014:251). We are bound by natural laws, and therefore, there are an infinite number of things that we simply cannot do, such as flying unaided, lifting a building from its foundations with our own strength or breathing underwater without diving apparatus of some kind. Secondly, they argue that even if a compulsory programme of moral bioenhancement is a definitive violation of our freedom of choice, it is not a self-evident conclusion that limiting our freedom of choice is always an entirely undesirable endeavour (ibid.). Our freedom of choice, regarding the range of possible actions we can perform, is impinged upon not only by natural laws, but also by societal laws, and to a certain extent by cultural norms – if the sanctions for their violation are sufficiently punitive so as to dissuade us from performing them. They therefore conclude that given a compelling enough justification, compulsory moral bioenhancement could be morally defensible.

The second potential response to concerns regarding freedom of choice would be to argue that if moral bioenhancement is voluntary then all freedom-related concerns, both freedom of choice and moral autonomy, are nullified. Thirdly, one could argue that even in the case of voluntary moral bioenhancement, freedom is at risk in some way. The risk to freedom that concerns the third

response could either be associated with a threat to moral autonomy, mentioned above, or, it could be associated with covert threats to freedom of choice produced by a number of other factors. Regarding covert threats to freedom of choice, a number of thinkers have discussed how incentivising moral bioenhancement could assist in persuading individuals to undergo it. Rakić argues that whilst state enforcement of moral bioenhancement would constitute a clear violation of freedom of choice, the state could provide "a variety of incentives in favour of morally enhanced citizens: tax reductions, schooling allowances for their children, retirement benefits and affirmative action policies that favour them" (2014:249)[27]. In other words, various social advantages could be used as a supposedly non-freedom subverting mechanism of persuasion.

However, to Rakić's claim that a lack of coercion implies freedom of choice, Selgelid argues that "[f]reely chosen moral enhancement would…not necessarily make moral enhancement compatible with freedom" (2014:215). This is because freedom of choice is more complex than it appears to be. Firstly, there are situations in which an initial freely-made choice can lead to impacts upon freedom. One such example is that of a drug-user who freely opts to utilise drugs, thereby becoming addicted and experiencing a loss of freedom to stop his drug use. As pointed out by Selgelid, one would generally not view drug addiction, or any addiction for that matter, as indicative of a state of freedom, regardless of whether or not the decision to take drugs was based upon an initial freely made choice (2014:215). This is a matter that will be discussed in detail in chapter 5. There is also a less obvious way in which freedom of choice could be impacted upon. In the case of providing incentives for moral bioenhancement, Selgelid points out that the persuasive pressure on individuals of not wishing to lose out on such incentives could be regarded as coercive. He posits that "the greater the costs of not doing something, the less free we are to do otherwise. [In the case of incentivising moral bioenhancement], foregone rewards count as costs" (Selgelid, 2014:216).

A potential application of moral bioenhancement, that supports the point that Selgelid makes, is the possibility of utilising compulsory – or voluntary - moral bioenhancement to 'treat' criminals. This possibility is receiving growing attention in the literature (Shook, 2012; Selgelid, 2014; Douglas, 2014; Curtis, 2012; Wiseman, 2014; Beck, 2015, Caouette, 2015). If aberrant behaviour

[27] Of course, it must be pointed out that if moral bioenhancement is justified on instrumental grounds, such as it being a plausible solution to the risk of ultimate harm, then such incentives will still not suffice as a means of encouraging universal adherence. Individuals with nefarious intentions will presumably not be swayed by such incentives and thus the threat will remain. Therefore, the moral justification for a project of moral bioenhancement would have to be supplemented by other compelling arguments such as the premise that it aims at enhancing an intrinsic good, of which the moral improvement of humanity in general would be one such example.

could be linked to an underlying pathology such as anti-social personality disorder, for example, then this would, strictly speaking, be a form of treatment rather than enhancement. Shook suggests the use of "restorative moral enhancers" as a possibility for "repeat offender[s]" (2012:10) and equates it with the way in which we currently treat individuals suffering from mental illness in order to enable them to function as members of society. He argues that the use of such enhancers to raise "conduct to minimally expected levels of civility" (Shook, 2012:10.) would, in all likelihood, receive public support. Shook posits that such a project could be justified on the basis of an appeal to a "model of intercultural objectivism" (ibid.). In other words, the use of moral enhancers could not simply be used to enforce specific cultural idiosyncrasies, but could only be justified if used to address behaviours and actions deemed universally problematic or pathological.

Selgelid suggests that in certain contexts, compulsory moral bioenhancement could result in "net gain[s] of freedom for those coerced" (2014:215). In the cases of addicts and criminals, their freedom is already compromised in some way: the former by the physical nature of their addictions and the latter by the fact that their behaviour has a strong likelihood of leading to their being apprehended and incarcerated. In this regard, Selgelid argues that "the freedom enabled by law-abiding life might thus outweigh the freedom costs of mandatory intervention" (2014:215). With regard to the threat to the freedom of criminals subjected to moral 'treatment', Douglas argues that "committing a crime might render one morally liable to certain forms of medical intervention" (2014:101). Chemical castration is one such intervention that is commonly practiced in the case of individuals found guilty of paedophilia. In such cases, freedom of choice regarding whether or not to undergo such a treatment is upheld, to a certain extent, as the offender must agree – or not – to the administration of a "medical corrective" (2014:103) as a requirement of early release or parole. Whilst this appears to better uphold freedom of choice the presence of coercion nevertheless remains, for similar reasons raised by Selgelid above.

There is yet another concern regarding the implementation of moral bioenhancement that has been explored by Crutchfield. He argues that moral bioenhancement is at risk of problems related to implementation due to "epistemological difficulties" (Crutchfield, 2016:390). While moral bioenhancement could be administered voluntarily or in a compulsory manner, where everyone is aware that it is taking place, Crutchfield explores a third possibility and develops an interesting argument to illustrate why the efficacy of any moral bioenhancement programme will only be ensured if it is implemented uniformly, but without public awareness. As discussed in the previous chapters, there are different targets of moral bioenhancement. Crutchfield posits that the two most

obvious targets of a moral bioenhancement programme would be either the "representational mental states" (2016:390) of individuals – with the most likely target being their *beliefs* – or, their "non-representational states" which includes their "affective states…[with the most likely target being their] motivations" (Crutchfield, 2016:390). The former, representational states have cognitive content whilst the latter do not. One could then 'change' the moral beliefs of an individual by administering, for example, oxytocin, which would then increase her empathy levels, thus resulting in her coming to have a belief that one should help those who are less fortunate. (Crutchfield, 2016:391).

The problem that Crutchfield foresees, however, is that people would realise that there had been a change in their beliefs, as they would remember their previous beliefs. What's more, he argues that they would realise that the change could be attributed to the administration of oxytocin rather than being the product of their own deliberation, which would, he posits, make them less likely to act upon these new beliefs. Crutchfield also foresees an additional problem that could make an individual less likely to act upon new beliefs. If an individual knew that he had been subjected to moral bioenhancement, and, in any way, doubted the ethical expertise of those enforcing such a programme, this doubt could take the form of "an epistemic virus that infects a person's moral psychology" (Crutchfield, 2016:396). This lack of trust could then have the result of persuading the individual to explicitly resist what he feels compelled to do. In order to ensure that individuals would assimilate and act upon their new beliefs, moral bioenhancement would therefore, Crutchfield argues, have to be administered in a covert manner. As suggested by Persson and Savulescu elsewhere, this could be achieved in the same way that fluoride is administered in drinking water (2008:174). Here, Crutchfield is suggesting that people would be more likely to assimilate new beliefs as legitimate motives for decision-making if they believe these beliefs have originated from their own volitions, even in the absence of a justification that can explain why their beliefs have changed.

Selgelid, however, has argued that focusing the debate on whether or not moral bioenhancement should be compulsory or voluntary obfuscates the issue to a certain extent in that it implies that we are left with one of only two choices (2014:216). This opposition is difficult to resolve and thus leads to a stagnation of the debate. Rather, it may be helpful if we look at the issue of implementation in terms of "degrees of encouragement…[and] degrees of discouragement" (ibid.). The degree of encouragement refers to the persuasiveness of the rewards for partaking in moral bioenhancement, whereas the degree of discouragement could refer to the persuasiveness of the

negative implications associated with not partaking in moral bioenhancement. For example, there could be various reward mechanisms associated with opting to morally bioenhance as opposed to a variety of punitive mechanisms associated with not doing so. Reframing the debate in this way would assist in transforming our view of freedom from an absolutist view – either we fully have it or we don't – to perceiving it as something that occurs on a continuum, and, which is thus possessed in degrees (Selgelid, 2014:216). In terms of this conception, Selgelid points out that at the far ends of both sides of the spectrum, freedom would risk being compromised. In other words, both strength of encouragement and strength of discouragement are linked to the perception of disadvantage of refusal. Put another way, the more compelling the encouragement to partake in moral bioenhancement, and thus, the more individuals feel they will be disadvantaged by missing out on incentives through not partaking, the more their freedom will be compromised. At the other end of the spectrum, the more compelling the discouragement of failing to partake in moral bioenhancement, and thus, the more individuals feel they will be disadvantaged by punitive measures, the more their freedom will be compromised. Reconceiving the issue of implementation in these terms rather than in absolute terms may be useful in moving beyond the deadlock that currently characterises the debate.

3.3.2 Administration problems

In addition to concerns regarding implementation, there are other concerns that pertain to the administration of moral bioenhancement. One such concern, that can be understood as *The problem of who will guard the guardians*, is encapsulated by a collection of related questions that would come to the fore if the science of moral bioenhancement became a possibility and the ethical concerns associated with it were resolved in favour of embarking upon such an endeavour. Questions that must be addressed in this regard are: who will decide what are suitable targets for moral bioenhancement; who will oversee this process and how may the project be protected from abuse, exploitation or nefarious intentions such as the furthering of moral eugenics agendas or behaviour control? This concern is strongly connected with what Persson and Savulescu describe as *The bootstrapping problem* (2012:2). This refers to the problem that it is morally deficient, or, at least, morally unenhanced human beings, the targets of moral bioenhancement, who must oversee the process of moral bioenhancement and ensure that it occurs in an ethical manner (Persson & Savulescu, 2012:2-3). Such concerns have received scant attention in the literature for a number of possible reasons. The moral bioenhancement debate is a relatively new one given that the relevant science is still at a rudimentary level. As a result, most discussions are still focused upon the ethical status of moral bioenhancement – namely, whether or not it is morally justifiable

– rather than on logistical problems and ethical concerns regarding how it would be administered. Due to the fact that most commentators in the debate are opposed to moral bioenhancement, it is understandable that less attention has been paid to administrative issues – or in-practice objections – than to in-principle objections. To a certain extent, a discussion of how moral bioenhancement would be administered may be seen as premature, as it presupposes that all the scientific and ethical concerns would have been resolved, so as to have arrived at a point where practicalities would need to be addressed. Nevertheless, it remains an ethical concern that warrants attention in the literature, particularly by the advocates of moral bioenhancement.

Persson and Savulescu provide various arguments in support of mandatory moral bioenhancement as an accompaniment to cognitive enhancement (2012:2). Their concern is that cognitive enhancement will enable individuals to acquire more extensive knowledge of dangerous technologies that could place humanity at risk for ultimate harm. All that is required to produce such catastrophic harm is the presence of ill-intentions, sufficient motivation to act upon these intentions, and, of course, the technological ability to carry these intentions through. Due to the fact that cognitive enhancement would vastly increase the cohort of individuals who possess the relevant technological abilities in this regard, Persson and Savulescu argue that "only beings who are morally enlightened, and adequately informed about the relevant facts, should be entrusted with such formidable technological powers as we now possess" (2012:2).

There are a few thinkers who have discussed the above-mentioned concerns, albeit, in a somewhat perfunctory manner. Murphy has pointed out that Persson and Savulescu, with the above-mentioned statement, imply that as technological ability increases, so too should positions of power be occupied by morally enhanced individuals (2015:375). However, the bootstrapping problem is particularly apparent here, regarding how this could actually happen. Namely, who, or which parties, would ensure that the individuals occupying such positions of authority abided by such a requirement? Beck has reframed the problem of who will guard the guardians by asking the question in terms of "who is likely to profit from moral enhancement and has an interest in its implementation?" (2015:238). Beck is, however, asking these questions with a different focus to the one I have in mind. She argues that the practical benefits of moral bioenhancement for individuals would have to be explicit enough to warrant undergoing such an intervention. While Persson and Savulescu emphasise the existence of an urgent need for moral bioenhancement, there has to be public agreement with them that is sufficiently persuasive. For the focus of this section,

however, Beck's question would require answering should a programme of moral bioenhancement ever become a possibility.

Arnhart has also posed the question of "who would be responsible for setting and enforcing the standards for [any] virtues and vices" (2010:80) that are taken as the goals of moral bioenhancement. He argues that the most obvious candidate would be the state. However, there would presumably be little public support for state-driven moral bioenhancement due to previous atrocities associated with state-controlled eugenics agendas and general fears of a Huxleyan type outcome[28]. Another option that Arnhart explores would be to have a variety of possible moral bioenhancements available, from which individuals could then choose. This could include permitting parents to choose particular virtues for their children through genetic interventions, a possibility that would, of course, elicit a host of ethical concerns that would require resolution[29].

In a debate between Savulescu and Harris, the above issue is directly addressed (2015). In answer to the question of who will guard the guardians, Savulescu points out that we are already in a situation where we have guardians who have tremendous power. For example, decisions such as which actions are punished and how greatly these punitive measures will impact upon personal freedom are legislated and enforced by all democracies through public consent. In other words, the public who elect governments through the process of democracy are the guards of the guardians. Therefore, they argue that there is no reason to suppose that this would change with a programme of moral bioenhancement (Savulescu, 2015:18). In addition, parents are given the freedom to decide how best to morally educate their offspring and the sum total of these freely made decisions bears a large contribution to how the moral standards that characterise a society are produced. Furthermore, regarding who will decide upon the nature or targets of moral bioenhancement, Savulescu posits that the obvious candidates would be those values and qualities for which there is

[28] Here I refer to Aldous Huxley's dystopic novel *Brave New World* (1932) in which the world state is led by ten world controllers who oversee the administration of the drug Soma that produces an effect of happiness, well-being and contentment, thus ensuring population wide acquiescence through keeping problematic emotions at bay.

[29] A number of thinkers have discussed the issue of moral bioenhancement via genetic selection, a process that would entail parents having more control over the genomes of their offspring (Walker, 2009, 2010; Murphy, 2015; Persson and Savulescu, 2015a; Faust, 2008, Arnhart, 2010; Blackford, 2010). At its most rudimentary level, this would entail a form of preimplantation genetic diagnosis whereby if it became possible to pinpoint particular genes that are associated with pathologies in psychological functioning, such as anti-social personality disorder, then only embryos that do not possess such genes would be implanted. This would not entail enhancement as such, because no changes would be made to selected embryos. At the more advanced stages, moral bioenhancement of embryos would entail modifying the genome of embryos to select for morally desirable dispositions. The ethical implications of parents being able to select the dispositions and abilities of their offspring has been discussed at length in the bioenhancement literature and is often framed as the 'designer baby' debate. It is, however, beyond the scope of this chapter, and this dissertation, to address this vast area of debate. For the purposes of this dissertation, I will narrow the focus to the ethical implications of freely chosen or universal moral bioenhancement for individuals themselves.

some form of universal consensus regarding their good. In particular, he argues that "we should aim for an ethic that promotes the values of justice, tolerance, respect for human equality, a sense of altruism, and willingness to cooperate and make small self-sacrifices for the benefit of others" (Savulescu, 2015:19). However, Harris responds to this by arguing that values should be "constantly revisable" (2015:19). In other words, whilst these may seem to be uncontroversially good values to choose, once they are biologically entrenched, or genetically selected, this puts their 'revisablity' at risk.

If Savulescu is correct, and the best answer to the question of who will decide upon the moral standards or values that could inform a project of moral bioenhancement is public agreement, then a project of moral bioenhancement would, of course, depend upon whether there are ethical standards that are universally recognised. If there is no consensus regarding the claim that there are certain values and standards that are interculturally objective or universal, then agreement regarding how moral bioenhancement could proceed would be impossible. In other words, one would have to give more credence to ethical objectivism as opposed to ethical relativism[30] to believe that moral bioenhancement is feasible. However, even if one were to subscribe to the former view, as Shook points out, there are, of course, competing theories within the ethical objectivist camp. Shook does not view this as signifying that consensus is unattainable, he posits that there is the possibility of an "overlapping consensus about some moral matters" (Shook, 2012:4). The fact that terms such as "'morality' and 'moral'…[are] meaningful to people, albeit in diverse ways" (ibid.) is evidence of this.

Regarding the bootstrapping problem, Sparrow argues that beyond simply identifying that it exists, Persson and Savulescu, pay it no further attention (2014a:28). All they say on the matter is that there has to be sufficient desire for moral improvement, which has to translate into ensuring that the requisite research is able to take place; and furthermore, if this research comes to fruition, it must be administered in an ethical manner (Persson & Savulescu, 2012:124). The problem arises due to the fact that it is morally deficient individuals – humans in general, according to Persson and Savulescu – who must guide this entire process. Whilst they argue that caution is required,

[30] Shook has discussed this matter, and, in the context of its use in the moral bioenhancement debate, he defines ethical objectivism as the view that "the moral norms to apply should be the justifiably correct moral standards regardless of what any culture or individual happens to endorse" (2012:10). This view is contrasted with ethical relativism which posits that there are no universal ethical standards. Rather, conceptions of the good or bad are relative to different cultures or contexts.

they also imply that pessimism in this regard should not lead us to a blanket rejection of moral bioenhancement.

Sparrow, however, seems unconvinced and voices concerns regarding the risk for potential abuse that techniques of moral bioenhancement could introduce. The main concern that he has with moral bioenhancement is that "it implies that those people directing it know what being more moral consists in" (Sparrow, 2014a:29), and, in this regard, he views it as guilty of moral elitism. This is inevitable, given the fact that a decision regarding the nature of morality would have to be made in order to know what to morally bioenhance, and this would require selecting a particular account of morality. Furthermore, even if this decision was made through democratic procedures and reflected majority consensus, it is highly unlikely that there would be uniform consensus; and thus, those holding different views would be subject to the view of the majority. On the other hand, the nature of the bootstrapping problem is such that it is more likely that techniques of moral bioenhancement would be used by minority regimes, with nefarious agendas, to make populations more pliable (Sparrow, 2014b:28). Avoiding such a concern would require ballasting democracy and the presence of international regulation of such technologies which wouldn't address the former issue. This concern will be discussed further below.

3.4 Other possible social effects of moral bioenhancement

Similarly to the above-mentioned concerns, the concerns in this section, are also speculative in character as they are addressing possible practical implications of moral bioenhancement. Nevertheless, they warrant investigation in any discussion of the ethical status of moral bioenhancement that aims to be thorough. I will firstly examine the concern regarding distributive justice. This includes two opposing perspectives that are present in the literature. On the one hand, if moral bioenhancement, in terms of making individuals 'more moral', is viewed as beneficial, how may we ensure equitable access to the advantages that it may secure? On the other hand, if moral bioenhancement is viewed as burdensome, how do we ensure that this burden is not unfairly shouldered by the enhanced? Both perspectives seek to elucidate whether access or adherence to moral bioenhancement would exacerbate existing inequalities or create novel ones.

Secondly, I will examine the concern regarding the threat to egalitarianism that moral bioenhancement may pose. Egalitarianism is, of course, a multifarious term. However, it is generally predicated on the view that individuals have equal moral worth; and therefore, that this should be reflected in specified contexts. In its most basic or general sense it refers to the view that

individuals should have equality of access to rights, liberties and opportunities, of which political equality would be one such right. As Wilson succinctly points out, in the moral bioenhancement literature, the primary conception of egalitarianism is the view that "relations of fundamental political equality are good in themselves" (2014:35). The concern here is that moral bioenhancement will undermine this fundamental political equality in some way. The most likely threat seems to be the view that if moral bioenhancement results in some individuals who are of a higher moral calibre, then they may have reason to consider themselves more suited to positions as political decision-makers than the unenhanced. Not only would they have a stronger claim to such positions, but it would also be in the interests of the unenhanced to grant them this claim. In this way, moral bioenhancement could undermine egalitarianism.

3.4.1 Distributive justice concerns

A small number of thinkers have commented on the potential implications for distributive justice of moral bioenhancement. Distributive justice concerns are typically associated with bioenhancement in general, where the enhancements in question are viewed as beneficial for the individuals who undergo them. Any intervention, positive state of affairs, or 'good' that is deemed to be advantageous for an individual is generally unanimously desired. When only certain individuals have access to such advantages, then we are faced with a distributive justice issue. Distributive justice concerns are, therefore, not specific to enhancement; they pertain to most institutional benefits and to resources in general. Thus, in the case of distributive justice concerns regarding bioenhancement, one could regard such concerns as symptomatic of deeper societal and global problems that require addressing in an arena that is distinct from that of bioenhancement. Nevertheless, it is useful to investigate how these concerns are addressed in the literature due to the fact that they may shed further light on moral bioenhancement itself. In the case of moral bioenhancement, the concern for distributive justice is of a unique nature.

As mentioned above, most bioenhancements – providing their safety has been established, and any potential ethical concerns have been resolved – would be regarded as unequivocally good for the individual receiving them. Some enhancements are viewed as positional goods in that the advantages they secure are dependent upon others not possessing this ability. Athletic abilities are an example of a positional good, because in order to win a race one must be faster than one's competitors. Tallness, which in particular societies is viewed as physically attractive or commanding, is also often discussed as an example of a positional good. This is because tallness is a relative notion; an individual only appears tall in comparison to others who are not as tall.

Thus, if the possibility arose of somehow genetically selecting for tall offspring, their ensuing tallness would cease to be advantageous if everyone were able to select for this characteristic in their offspring.

Some enhancements may be considered to be both positional and non-positional goods. Cognitive enhancement is one such example. Improving one's cognition will, on the one hand, confer personal, positional advantages in competitive environments. On the other hand, if cognitive improvements are utilised to increase one's performance in positions that are beneficial to the public in some way, then this would be an example of a non-positional good. For example, if a cognitive enhancement enables a scientist to discover a cure for cancer then this enhancement will be beneficial for everyone, rather than simply for the scientist who discovers it. In the case of a non-positional good, its goodness is not lessened the more extensively it is shared or held, whereas this is the case with positional goods, to a certain extent. General bioenhancements, such as improvements in human physical and mental functioning, are vulnerable to the criticism that because they are mostly positional goods that will benefit only the individual, possibly at the expense of others, they represent selfish or greedy aspirations.

Most forms of general bioenhancement are therefore subject to distributive justice concerns due to the fact that if access to the benefits they confer is impeded in any way then this constitutes an injustice. Moral bioenhancement is, however, clearly not a positional good, and is furthermore unique in that the benefits it will confer are primarily public benefits. As Douglas points out, rather than placing others who are not morally enhanced at a disadvantage, the opposite will be the case with morally bioenhanced individuals (2008:230). For some, this is regarded as problematic. Bronstein, for example, remarks that the kinds of interventions discussed in the moral bioenhancement debate are aimed at securing societal benefits rather than individual benefits (2010:85). Murphy makes a similar point. However, he posits that whilst it may be true that the most obvious benefits associated with moral bioenhancement are public rather than personal, the types of enhancements that Persson and Savulescu are suggesting will not be detrimental to individuals (Murphy, 2015:271). It is difficult to see how an enhancement of empathy, for example, would lead to "inherent…[or] intrinsic disadvantage" (ibid.) for the individual in question. In other words, an individual would not necessarily incur personal psychological harm or be put at a "social disadvantage" (ibid.) by caring more about the welfare of others.

This is, however, not to say that there are no potential advantages for the individual that may be associated with moral bioenhancement. An Aristotelian view, for example, equates the acquisition of virtues with human flourishing (Aristotle, 2004). In other words, to lead a 'good' life in which one flourishes as a human being and achieves a state of *eudaimonia*[31], one must constantly 'practise' being good. Thus, undergoing moral bioenhancement could be associated with personal benefits for the individual if it enables her to lead a happier, more fulfilled life than she otherwise would have. The claim is not that moral bioenhancement would create an artificial happiness, but rather, that one is more likely to experience happiness and fulfilment from being morally upright than morally deficient.

Focquaert and Schermer have also remarked that moral bioenhancement is viewed as problematic for a number of reasons, one of which is that "the advantages of moral enhancement may fall upon society rather than on those who are enhanced" (2015a:140). To the extent that moral bioenhancement is able to attenuate behaviours that impact negatively upon society, such as criminality, this is true. However, in such cases, it would be odd to argue that individuals who no longer desire to take part in criminal activities are being personally disadvantaged in some way. Focquaert and Schermer's claims are perhaps more relevant in cases where moral bioenhancement encourages selfless or supererogatory behaviour. For example, if an individual's altruism and sense of justice are enhanced, thereby producing distress at the suffering of others, which in turn compels him to make personal sacrifices or to take action at his own expense, then this could be a case of an enhancement that would be beneficial for others but not for the individual in question. However, if more individuals were sufficiently compelled to take action to secure the welfare of others less fortunate than themselves, at acceptable levels, this would be an argument in favour of moral bioenhancement as a means of serving the ends of distributive justice.

Casal examines the above concern and argues that if we could establish "an appropriate threshold of compliance...to moral norms" (2015:341) and then utilise moral bioenhancement to elevate individual behaviour that falls below this threshold, then this could provide a legitimate argument for moral bioenhancement as increasing justice. In other words, behaviour that falls below this threshold – which would include criminality, free-riding, or generally selfish behaviours that have a negative impact upon society – places an unfair burden on those whose behaviour falls above the threshold and is therefore unjust. However, as Casal observes, regardless of whether or not one

[31] This Greek term is not easily translatable. It is sometimes simplified to signify happiness; however, most argue that its meaning is more akin to a state of flourishing or the sense of well-being that comes with fulfilling one's purpose as a human being.

views "compliance with morality to be a burden or a benefit" (2015:342), there are justice based arguments that can be made from both perspectives. If one holds the view that one is disadvantaged by respecting moral norms, as well as the law in general, if there are others who do not reciprocate, then moral bioenhancement could be viewed as a means of addressing this disadvantage or burden (ibid.). If, on the other hand, one sees moral acuity as advantageous and moral deficits or criminal tendencies as disadvantageous, then preventing treatment which may correct these deficits could be viewed as unjust for those who are morally less than exemplary.

Morioka also looks at the distribution of benefits and burdens regarding moral bioenhancement and argues that if mandatory moral bioenhancement were somehow implemented, immoral individuals and those with financial and social resources would be more likely to evade it (2014:122). In this way, it is likely that moral bioenhancement could result in two distinct classes of individuals. Furthermore, there is a strong possibility that morally bioenhanced individuals would be vulnerable to abuse and exploitation at the hands of the unenhanced group. The nature of the moral bioenhancements that are considered by Persson and Savulescu, such as those that increase trust and altruism, are such, that this abuse and exploitation may be met with little resistance.

Beauchamp considers the opposite possibility, namely, that moral bioenhancement could "exacerbate, rather than diminish existing social prejudices and distributive unfairness" (2015:347) due to initially only being available to the affluent. In the same way that the affluent are able to secure the best education and resources for their children, so too, could they ensure that their children are not afflicted with "negative dispositions" (Beauchamp, 2015:348). In this way, those who do not have "access to bioenhancement techniques will likely be rendered worse off, relative to the more advantaged, than they now are" (ibid.). Implicit in Beauchamp's argument here is the above-mentioned view that moral acuity equates with leading a flourishing life and can thus be seen as beneficial. If this is correct, then, in the interests of distributive justice, moral bioenhancement would have to be available to all. However, as Beauchamp points out, human moral deficiencies, as outlined by Persson and Savulescu, give us cause for concern regarding the likelihood of this (ibid.). Furthermore, it is likely that inequalities in access will be experienced not just within nations and amongst individuals, but predominantly between affluent and less affluent nations

It is interesting to the note the diversity in responses to the issue of whether the enhancement of morality would be regarded as beneficial or burdensome for individuals. The view that being more

moral would place one at a personal disadvantage is highly concerning and is perhaps evidence in support of Persson and Savulescu's claims that human beings in their current state are morally deficient. This view seems to originate in the perception that moral decency, or at least attributes such as altruism, fairness and trust, may render one more easily exploitable. However, this is also possibly based on a misconstrual of how moral bioenhancement would work. As Walker points out, it seems to imply that "the virtuous are meek or compliant" (2009:42), a view which is unsubstantiated. To support his claim, Walker discusses various examples of moral heroes, such as Gandhi, who sought political transformation through peaceful mechanisms but could never be viewed as docile or subservient (2009:43)

3.4.2 Egalitarianism versus moral perfectionism

The issue of the implications for egalitarianism of moral bioenhancement elicited spirited ethical debate in a particular special edition of the *American Journal of Bioethics*. This edition presented a target article by Sparrow and a number of responses to his claims (Sparrow, 2014; Wiseman, 2014; Robichaud, 2014; Jotterand, 2014; Lechner, 2014; Ram-Tikten, 2014; Rakić, 2014; Marshall, 2014; Wilson, 2014; Persson & Savulescu, 2014b). This set of articles is useful as most of the relevant concerns for egalitarianism are identified and addressed.

Sparrow is not a supporter of moral bioenhancement and expresses scepticism firstly, as to whether the project, as envisaged by supporters such as Persson and Savulescu, would actually constitute an enhancement of morality and secondly, whether it would succeed in its stated aims (2014:20). However, for the purposes of his discussion, he assumes it could be successful and investigates what the consequences could be for egalitarianism. His main focus in this regard is the implications for egalitarianism that could arise due to the existence of a "class of citizens [that] might be, as a matter of biological constitution, morally better than another class of citizens" (Sparrow, 2014:22). Some of Sparrow's concerns for the impact of moral bioenhancement are connected with the problems discussed in the above sections, namely: on whose account of morality will a project of moral bioenhancement be based and how could it be implemented. Furthermore, the question is: if the project of moral bioenhancement did result in a class of morally superior individuals, should these individuals not have a greater contribution and impact upon public decision making (Sparrow, 2014:24)? This could lead to a society similar to Plato's *Republic*, in which a special class of citizens, the guardians, are granted ruling power based upon their superior intellectual and moral abilities (1985). However, this would only seem to be problematic if access to moral bioenhancement was not freely available to all. If the ability to become morally and intellectually

'superior', and thus, to be deemed suitable for public office, is freely available then any charge of elitism is severely weakened

The above concerns are associated with a successful project of moral bioenhancement; however, Sparrow's deeper concerns are directed towards another possible outcome: its failure (2014:26). Firstly, moral bioenhancement could fail in a way that its failure is obvious. Secondly, and of more concern, it could fail without us realising. In other words, we could erroneously think it had been successful. Regarding the second concern, as mentioned above, if there was a means of identifying whether or not moral bioenhancement had been successful this could lead to the biological categorisation of human beings into two groups with the successfully morally enhanced occupying positions associated with greater political privileges. There are good consequentialist arguments as to why this would be beneficial, all things considered. However, if the project failed in some way and its failure was not established, this could lead to the mistaken conferral of political privileges with potentially negative consequences. Furthermore, a scenario that Sparrow considers to be of far more concern, is the possibility of a small group of elites acquiring political power, with the support of the majority, on the basis of their purported successful moral enhancement, when in actual fact they are fully aware of the failure of their enhancement. In this regard, it could be an additional means of facilitating the illegitimate authority and tyranny of a small group.

Sparrow concludes his discussion by pointing out that while much of the moral bioenhancement debate is purely speculative and philosophical at this point; it may have serious practical implications (Sparrow, 2014:26). The most problematic outcome is that the continued debate concerning the possibility of biologically manipulating our morality could serve as a means of disseminating the notion that "some people…[are] naturally better people than others" (Sparrow, 2014:27). As mentioned in section 2.4.2, implicit in such a view, Sparrow argues, is the belief that whether one is moral or immoral, is a product of one's biology and therefore that then there is very little that can be done in this regard: "those who are immoral are incorrigibly so, while those who are most moral are good by nature" (ibid.)

Sparrow's concerns have elicited a number of responses. Persson and Savulescu point out that moral bioenhancement can be presented as either "a confident…[or] cautious proposal" (2014:39). They argue that most of the concerns raised by Sparrow are associated with the former type of proposal, whereas they posit that their argument should be viewed as an example of the latter. Their argument for moral bioenhancement asserts that it is a *possible* solution to the problems they

outline, that cognisance of risks and safety is paramount, but that it nevertheless merits further research. To Sparrow's claim that moral bioenhancement supports an agenda of "moral perfectionism" (2014:40), by virtue of the fact that it picks a particular account of morality, despite the presence of different interpretations of the good, Persson and Savulescu, once again, argue that we already do this when we educate our children morally through traditional mechanisms. They point out that their suggested potential targets for moral bioenhancement – the moral dispositions of reciprocity, or a sense of fairness, and compassion – are already viewed as uncontroversial goals of traditional moral education. Furthermore, their desire to target these dispositions as a means of addressing "free riding and criminal behaviour" (Persson and Savulescu, 2014:40) is a goal shared by most societies. Thus, they argue that if Sparrow is to argue that moral bioenhancement is problematic, due to the fact that it presupposes a specific interpretation of moral goodness, then he must explain why it does this to a greater degree than traditional mechanisms of moral enhancement

To Sparrow's concern that moral bioenhancement could lead to two morally differentiated classes of human beings, Persson and Savulescu point out that it is already the case that there are clear differences between individuals regarding their moral behaviour (2014:41). Other thinkers have also emphasised this point (Marshall, 2014; Jotterand, 2014; Robichaud, 2014). However, this has not led to the formation of distinct classes based upon the moral predilections of individuals. In fact, as they argue, the biological components of human abilities are, in general, characterised by great inequalities regarding how they are distributed. Therefore, a clear argument could be made that moral bioenhancement, and bioenhancement in general, would reduce inequalities rather than exacerbate them. In this regard, it could be viewed as supportive of egalitarianism.

Marshall also addresses this concern (2014). She separates Sparrow's claims and addresses each one in turn. Sparrow argues that society is characterised by a diversity of moral responses; there are some individuals who are clearly more moral than others. He posits further that it is unlikely that moral bioenhancement would produce the same effects for all individuals. In the same way that individuals react differently to the same medications or medical procedures, moral bioenhancement could further boost those who are already morally praiseworthy and possibly have minimal effects on less moral individuals. If political privileges are then awarded on the basis of optimal morality, then this could be unfair as the more moral will then secure further benefits at the expense of the lesser moral.

In response to this set of claims, Marshall argues that it is "plausible but not inevitable" (2014:29) that moral bioenhancement would further boost existing moral preclusions further. However, similarly to Persson and Savulescu, she posits that it is equally as likely that it could produce the opposite effect. It could give rise to the most acute improvements in the morality of those whose moral attributes are weaker, thereby raising average levels of morality and lessening the gap that concerns Sparrow (Marshall, 2014:29). Sparrow's concerns regarding negative outcomes are, of course, speculative claims that would have to be settled by empirical means. Furthermore, as Marshall points out, even if Sparrow's concern were to be realised, it is still not evident that 'more moral' individuals would be the only ones to enjoy the ensuing advantages (ibid.). Presumably the benefits reaped by morally competent individuals who possess the political power to successfully address the problems that concern Savulescu and Persson would be enjoyed by all, regardless of political office. The only personal benefit exclusive to those who would be part of the political process would be "some small benefit that comes from enjoying being part of the collaborative process itself" (Marshall, 2014:29). However, regarding Sparrow's general concern pertaining to the development of two separate moral classes of individuals, Marshall argues that this could produce unanticipated "social and political problems" (2014:30) and therefore we should pay adequate attention not only to the purported advantages that moral bioenhancement may produce, but also to the potential problems that may ensue.

Regarding the above potential threat to egalitarianism posed by a class of distinctly morally enhanced individuals who could then be regarded as potentially more deserving of certain political privileges, Robichaud argues that Sparrow's concern in this regard is misdirected (2014:33). He argues that even if we grant to Sparrow the possibility that this could occur, the danger for egalitarianism would be posed not by moral bioenhancement itself, but rather, by "the ability reliably to pick out agents who are more sensitive and reactive to moral reasons" (Robichaud, 2014:33). Robichaud argues that Sparrow mistakenly attributes this ability to moral bioenhancement. In democratic societies, it is generally viewed as preferable, to vote into office – thereby conferring certain benefits and powers – individuals who are perceived to possess moral integrity. This is based upon the view that such individuals are believed to be more likely to act in the best interests of the public than individuals of dubious moral integrity. Of course, there are other capabilities that are also regarded as paramount, such as possessing the requisite knowledge, training and competency for such positions. In general, decisions regarding the moral integrity of potential candidates for office are made through various inferences and calculations, personal interpretations of behaviour and various other forms of evidence, such as how candidates are

presented in the media, and so forth. Robichaud's argument implies that it is the validity of this process of decision-making, and the information that supports or thwarts it, that produces the threat to egalitarianism that concerns Robichaud. If certain fields of research, such as neuroscience for example, were able to provide us with more sophisticated, and, most importantly, more *reliable* means of accurately discerning between the moral integrity of individuals, then it would be more appropriate to question the threat posed to egalitarianism by this mechanism than by the actual process that increases moral integrity (Robichaud, 2014:34).

Wilson takes a different approach to Sparrow's concern for egalitarianism. He points out that if moral bioenhancement were truly successful, these fears would be unfounded (Wilson, 2014). The specific fear that he is referring to here is that the morally enhanced – due to their possessing greater moral competence – would desire to exclude the non-enhanced from public decision-making arenas. However, if a project of moral bioenhancement included improving capacities of fairness or sense of justice and altruism, as envisaged by Persson and Savulescu, it is difficult to imagine that this would be so. The desire to exclude others from such areas is more likely to be evident in individuals seeking to further their own interests and power – in other words in the morally unenhanced – than in those who have been morally enhanced. More specifically, as Wilson points out, "we can expect the morally enhanced to possess a particular motivational set that is incompatible with harmful exclusion" (2014:35). Furthermore, if, as Sparrow argues, "the aim of democratic decision-making is to increase the probability of reaching the right answer" (2014:25), then it is not clear that morally enhanced individuals would have a justified monopoly in this regard. Such decision making would require moral as well as cognitive components and while morally enhanced individuals would possess the former, without having received cognitive enhancement they would be no more likely to arrive at "the right answer" than a morally unenhanced individual. Wilson posits that "competence in leadership and moral competence are two different things" (2014:36)[32]

A slightly different problem raised by Sparrow, that was mentioned above, is that moral bioenhancement discussions tend to reinforce the idea that goodness and badness are inherent to individuals, and thus, that there is very little one can do to change this (2014:27). A number of thinkers have responded to this claim. Firstly, Persson and Savulescu respond to this by arguing that they have never claimed, as Sparrow implies, that morality is entirely biologically determined.

[32] This then introduces the question of how effective moral bioenhancement would be in the absence of cognitive enhancement.

Rather, their argument, based upon research conducted on identical twins, is that the split between biological and environmental influences is roughly even (Persson and Savulescu, 2014:42).

Robichaud also responds to this claim. He argues that when making an assessment regarding the moral character of an individual, we generally take a number of factors into consideration. Such assessments are not only informed by our perception of their moral attributes and behaviour but also by the amount of effort they display in this regard and other important contextual factors such as their "personal history, or social roles" (Robichaud, 2014:34). Robichaud provides some examples here. Let us imagine an individual who is in prison, and appears to have suboptimal moral attributes or functioning. If this individual were to then come to the realisation that he has moral deficiencies and desires to address this; and then, through great effort, is able to change his behaviour, this would be regarded as morally virtuous. In addition, we regularly make concessions for those individuals whose moral development has been thwarted by factors in their personal histories. In such situations, we understand that certain undesirable patterns of behaviour are a product of the life experiences of the individual in question, rather than viewing them as indicative of inherent 'badness'. This indicates that there is an awareness of the:

> distance [that] exists between an agent's neurochemical and genetic endowment…[and that] there is no necessary conceptual link between the claim that moral capacities are biologically based and the highly dubious and elitist claim that the quality of an agent's character rests solely on her (enhanced or not) biological constitution" (Robichaud, 2014:34).

Sparrow does not explicitly state this, but perhaps the tacit concern is that if the impression is given that individuals are 'naturally' good or bad, this could then support an interpretation that the more moral amongst us are, in some way, more valuable, or possess greater moral worth. Therefore, enhancement, whether voluntary or not, could lead to a class of individuals who are considered to be of a higher "moral status", more valuable or worthy by way of their enhancement, than those who have chosen to remain unenhanced. This would then have implications for Sparrow's concern regarding the development of two distinct moral classes of individuals, if classes came to be viewed not only as distinct in kind, but also distinct in terms of their moral status or worth. In response to this, Jotterand argues that it does not follow from the fact that an individual acts in a more morally worthy manner, that he or she then has more moral worth than an individual who is less morally praiseworthy (2014:2). He argues that Sparrow commits a category error in his confusing "difference in kind (moral status)…[with] difference in degree (moral behaviour)" (Jotterand, 2014:2). We are already faced with a reality in which there is great diversity in the distribution of moral attributes; some individuals are clearly morally 'better' than others. However, this has not led us to utilise this state of affairs as a means of supporting a more extensive set of rights or

benefits for such individuals. Individuals are generally viewed as possessing equal moral worth on the basis of the intrinsic value associated with having personhood. It seems uncontroversial to state that – over and above the specific attributes and capacities viewed as appropriate and desirable to specific contexts – most individuals would prefer their personal interactions to be with individuals who are honest, compassionate and fair. However, it would be absurd to generalise from the basis that because it is more pleasant to be in the company of such individuals, they are therefore of a higher status or possess more intrinsic worth as human beings.

Many of the responses to Sparrow's concerns regarding the threat to egalitarianism argue that most of his concerns pertain primarily to compulsory moral bioenhancement. As discussed in section 3.3.1, most thinkers in the debate seem to hold the view that compulsory moral bioenhancement would be morally abhorrent. Thus, in terms of Sparrow's concern that moral bioenhancement would lead to a state-driven ethos of "controversial moral perfectionism" (2014:20), Rakić responds by pointing out that this would be circumvented if moral bioenhancement were voluntary (2014:37). If the decision of whether or not to morally bioenhance oneself is autonomously made by the individual, then the state cannot be charged with furthering such an agenda.

Ram-Tikten also responds to Sparrow's concern that in deciding upon what moral qualities would be targets of moral bioenhancement, the state, or whoever is part of this decision-making process, would be guilty of a form of "moral perfectionism" (2014:27) which would be at odds with the recognition of moral pluralism and egalitarian principles in general. However, Ram-Tikten argues that it is not "mere pluralism but *reasonable* pluralism" (2014:43), as espoused by Rawls, that could be our guide in this decision-making process. When deciding upon which values should inform the legislation of a society characterised by diverse views regarding the nature of 'the good', it is not the case that simply because a specific moral conviction or perspective exists, it therefore merits consideration. Rather, it is only "reasonable" opinions, or those that can be mutually agreed upon by a group characterised by different and possibly competing moral beliefs, that are considered. In other words, we could adapt Rawlsian political ideals for use in the moral sphere by only considering the moral attributes for which there is "overlapping consensus" (Ram-Tikten, 2014:43) as suitable candidates for moral bioenhancement. In this way, Sparrow's egalitarian concerns could be allayed. Brooks also explores the way in which reasonable pluralism could be helpful in identifying the targets of moral bioenhancement. He argues that only if moral bioenhancement alters behaviour in such a way that individuals are less likely to be able to consider the arguments for various reasonable doctrines will it pose a threat to egalitarianism (Brooks, 2012:29).

3.5 Concluding remarks

In this chapter, I have addressed the most prevalent in-practice objections to moral bioenhancement that appear in the literature. In-practice objections engage with the potential real-world risks, harms and benefits of moral bioenhancement, and are therefore, invariably consequentialist in nature. In the moral bioenhancement literature, such objections are typically provided in response to arguments given in support of moral bioenhancement and aim to show why the purported benefits will either not be realised or, more specifically, why moral bioenhancement will result in individual or general harms with practical implications. Furthermore, in-practice objections are either directed towards potential harms that may occur at an individual level, or, at societal level. The two areas of concern are entirely different in substance. At the individual level, risks are more concerned with potential physiological harms, whereas at the societal level, potential harm would be socio-political in nature.

Thus, the concern for risks and potential harms that could affect individuals is primarily an issue that would have to be resolved through scientific research, as no biomedical intervention would receive approval if it were not known to be safe. This would be even more true regarding the bioenhancement of 'normal' moral functioning. Whilst the safety and risk concerns for individuals of such interventions are inconclusive, the matter remains relatively ethically simple at this point, as the avoidance of harm is a cornerstone of biomedical ethics. In other words, safety concerns should be primary, and until it could be established that proposed interventions, as they develop and become possible, are safe for individuals, moral bioenhancement should be a matter of theoretical discussion only. Whilst addressing this concern is predominantly a scientific matter, ethicists will play a continuing role in elucidating why moral bioenhancement does or doesn't warrant the requisite research.

As mentioned above, the concern regarding potential harms and risks that could occur at societal or global level is of an entirely different nature, and therefore, there is much that the field of ethics can contribute here. In terms of the issue of how moral bioenhancement could be implemented, the matter seems irresolvable. To achieve what Persson and Savulescu demand of it – among other things, the aversion of ultimate harm – moral bioenhancement would have to be administered universally and possibly covertly. However, universal implementation would clearly be both practically impossible and ethically impermissible. Administering moral bioenhancement on a voluntary basis would avoid many of these problems but would be ineffectual in achieving

significant benefits. What the discussion of implementation illustrates, therefore, is that what is at stake are two freedoms: freedom of choice or action and the deeper freedom that is interpreted as moral autonomy. Violating freedom of choice would be impermissible, thus, the focus then turns to whether voluntary moral bioenhancement would violate moral autonomy.

The primary problem regarding the administration of moral bioenhancement would be related to resolving the problem of moral content in a way that respects moral pluralism and protects the public from any ill-intentions of those who would implement and oversee this process. Regarding the latter problem, this concern bears similarity to the trust we are currently required to place in leaders in the political arena who make decisions with far-reaching impacts upon our lives. Through the process of democracy, we elect the leaders who we believe will best serve and protect the public interest. In the case of the administration of moral bioenhancement there is no reason we would, and should not, continue to utilise the process of democracy. Concerning the problem of moral pluralism, the fact that there is already an overlapping consensus regarding what it means to morally educate our children in a desirable manner, is evidence that this problem may be overstated. Furthermore, I would argue that the moral dispositions of reciprocity, or a sense of fairness, and empathy for others, that have been selected by Persson and Savulescu as targets for moral bioenhancement come close to enjoying universal respect as morally exemplary qualities. Thus, it is not self-evident, as some thinkers claim, that moral bioenhancement is implicated in a form of moral elitism where the views of a minority regarding 'the good', will be foisted upon an unwilling and beleaguered majority.

Moral bioenhancement is unique in comparison with other forms of bioenhancement of human abilities as it would confer predominantly societal or group benefits, rather than individual benefits. Thus, while there is merit in the concern that those who are morally bioenhanced could be vulnerable to the unenhanced who could seek to exploit them, I would argue that this concern is overstated. This is because it implies that being moral makes one compliant, possibly naïve, and thus, easily exploitable; a view that isn't substantiated by empirical evidence. In other words, this concern seems to disregard the fact that moral bioenhancement would aim to bring individuals up to the levels of those who are regarded as moral exemplars within society, and not to a radical, never-before-witnessed level. Thus, to resolve this concern, we simply need to examine whether those that we regard as moral exemplars have been historically more vulnerable to exploitation.

Conversely, I would argue that the concern that the morally bioenhanced would demand greater political privileges at the expense of the unenhanced is misguided as it is founded upon a belief that morally bioenhanced individuals would desire to act in a harmful manner by excluding the enhanced which is incongruent with the aims of successful moral bioenhancement. If such a case arose, it would imply that moral bioenhancement had failed. Finally, an argument that requires more attention is the claim that if it could be shown that there are distinct personal advantages associated with possessing moral virtue, then withholding it from those who would wish to use it could be regarded as an injustice.

Chapter 4a – In-principle objections to moral bioenhancement: the concern for personal identity

4.1 Introduction and overview of chapter

As discussed in chapter 3, objections against moral bioenhancement can be made not only on the grounds that it has the potential to produce negative practical consequences – in other words that it is wrong in practice – but also on the grounds that it is morally objectionable in principle. The two types of objections are disparate to a certain extent, as something may be permissible in principle, or in theory, but not in practice; and vice versa[33]. As mentioned in chapter 3, in-practice objections generally deal with purported risks, benefits and harms, and are therefore primarily consequentialist in nature. In the case of moral bioenhancement, due to the fact that many of the proposed interventions are not yet possible, such objections rely largely on extrapolation from similar interventions or events, and are thus, predominantly speculative. Despite their speculative nature, however, their explication can be greatly assisted by drawing upon relevant empirical information where it is available. For example, physiological safety concerns regarding moral bioenhancement require input from scientific or medical fields in which similar procedures and interventions have been used for therapeutic purposes and may thus provide valuable information regarding the likely outcomes of such interventions when used for enhancement purposes. Political or social concerns regarding the potential impact of moral bioenhancement on a practical level could be investigated by drawing upon insights from a wide variety of disciplines within the social sciences. Explicating the status of in-principle objections, on the other hand, requires a different approach.

Due to the fact that in-principle objections are primarily philosophical in nature, their explication requires the tools of philosophy, namely, critical analysis, as well as rational deliberation and

[33] Uncontroversial examples of such situations are not easy to provide, as certain practices that would be regarded as permissible in principle by some, would be viewed as impermissible by others. Thus, these distinctions rely upon interpretation to a large extent. These difficulties aside, an example of an activity or practice that could be regarded as wrong in principle but permissible in practice would be torture. I can be opposed to torture, in principle, because I believe it to be intrinsically wrong to do such a thing to another human being, but condone it in practice. For example, in a situation in which an individual who has hidden a bomb – that will detonate imminently, resulting in widespread loss of life – refuses to reveal the whereabouts of the bomb, I may support the use of torture to force him to reveal the whereabouts of the bomb. Examples of situations in which something is permissible in principle but wrong in practice are easier to provide. War situations, in particular, provide such examples. In principle, if one is in the military and in a situation of combat, or war, it is generally viewed as permissible to kill one's opponents. This is particularly the case in situations of self-defence. While this may be permissible in principle, killing someone in practice could be a very different matter. A combatant could possess all the relevant facts such as the permissibility of killing in war, but the practical act of taking another life could be impossible for him. Another, more straightforward example, of the above would be a medical procedure that is ethically uncontroversial, but is impermissible in practice due to the magnitude of the risks involved.

argumentation. In the context of moral bioenhancement, opposing in-principle arguments generally claim that moral bioenhancement is intrinsically wrong, despite any positive outcomes it may produce. In other words, for various reasons, it is argued to be wrong, in and of itself. As in-principle objections to moral bioenhancement engage with notions of intrinsic or absolute wrongness, they are generally extremely difficult to reconcile with the kinds of – generally consequentialist – arguments provided in support of moral bioenhancement, such as Persson and Savulescu's argument. However, this impasse is characteristic of the gulf between consequentialist and non-consequentialist moral theories in general, and in particular, between the former and moral realism, and is therefore not specific to moral bioenhancement[34].

There is, however, some overlap between in-principle objections and in-practice objections. This is because whilst an in-principle objection may be predicated upon the claim that a proposed endeavour or intervention is wrong regardless of any positive effects it may produce, it is very seldom the case that this purported wrongness is not also connected to some real-world impact that the endeavour or intervention may produce. In this regard, other disciplines, such as the sciences, may be of assistance in clarifying in-principle objections. This is especially true of the most pervasive in-principle objection found in the literature, namely, the concern that moral bioenhancement will impact upon human moral autonomy and will therefore impair or eradicate morality itself. John Harris, and a number of other thinkers, are of the view that morality, and more specifically our moral autonomy or agency, is of absolute value; therefore, any interventions that threaten this, regardless of the benefits involved, are impermissible. His argument is, of course, informed by his view of what constitutes moral autonomy, and in this regard, such an objection may be explicated with information from empirical fields such as neuroscience and moral psychology in which the physiological and psychological underpinnings of such phenomena have been studied. This concern would be an example of the way in which empirical information would be of assistance in supporting or dispelling some of the assumptions that are being made by supporters and opponents of moral bioenhancement regarding in-principle objections.

This chapter is divided into two sections, with both addressing in-principle objections to moral bioenhancement that have appeared in the literature. In the first part, I will discuss one of the more neglected areas of the debate; the concern that moral bioenhancement may impact upon personal identity. The sense in which personal identity is interpreted in the context of the moral

[34] Moral realism is a meta-ethical perspective that argues that moral facts have objective, and thus, independent truth. A moral realist would argue that acts that are right or wrong, are right or wrong in a universal or absolute sense; that is, their rightness or wrongness is independent of any contextual facts or outcomes that they may produce.

bioenhancement debate will, or course, be discussed throughout this chapter as well as in chapter 5. It is important, however, to point out, at the outset of this chapter, that in the moral bioenhancement debate, personal identity is not being interpreted in a metaphysical sense[35]. Rather, the sense in which it is being interpreted here, is more qualitative or phenomenological in nature. What is at stake in the moral bioenhancement debate, in terms of the concern for personal identity, is captured by concern for what I take to be unique to myself; my sense of selfhood. This would refer to the nature of the structural relationship between my thoughts, beliefs, attitudes, desires and preferences that make me recognisable to myself, and to others, as a *specific* character or personality. In other words, my *particularity*. One may concede that there is a first-person *experience* of some core of the self that endures and is recognisable through time without being required to make further concessions, of a metaphysical nature, regarding the nature of this self.

In the second part of this chapter, I will then investigate the most pervasive in-principle objection to moral bioenhancement, the argument that moral bioenhancement will compromise moral autonomy, or agency, and thus, morality in general. I have kept these two objections in the same chapter, rather than presenting them in separate chapters, as I will argue in chapter 5 that the two concerns are related. If by personal identity, what is meant is our sense of selfhood or self-conception – those characteristics of the self that we believe constitute who we are, and, most importantly, that we identify positively with – then major unanticipated impacts produced by moral bioenhancement in this regard, *could* be viewed as compromising our autonomy, where the latter is understood as our sense of self-determination and authenticity. Of course, whether this would be the case would depend upon how attached we are to the self we perceive ourselves to be. Individuals are inured to constant changes to what they would regard as their self-conception. Generally, this is accepted as an aspect of life that is unavoidable, and perhaps even rewarding for some individuals. What I take to be morally salient regarding the concern for personal identity is the *intensity* and *source* of such changes, as well as the *attitude* of the individual in question towards such changes. Major, rapid or unwanted identity changes that arise due to biomedical interventions and those changes that produce a sense of inauthenticity regarding one's selfhood would be the kind of identity impact that would warrant caution.

[35] In this regard, a further caveat is necessary. I am, of course, aware that in terms of certain philosophical debates, identity is a highly contested notion that is interpreted in a variety of different ways. There is, for example, a rich history of debate regarding the nature of identity within the philosophy of mind. In addition, identity is also a notion that lends itself to utilisation in the service of a variety of aims and agendas, some of which are political or ideological in nature. While I take note of this, I will not be engaging with identity in these above-mentioned terms.

Of course, whether or not one regards moral bioenhancement as posing a risk to personal identity will depend largely on the conception of identity that one holds. Throughout section 4.2 I will therefore investigate a variety of interpretations of identity that have been discussed in the literature. The discussion will be informed by insights from the fields of neuroscience and psychopharmacology. In particular, the discussion will include examples of impacts on personal identity produced by therapeutic interventions utilised in these fields. Discussion in this regard is useful because some of these interventions have been proposed as mechanisms of moral bioenhancement in the literature.

In section 4.2.1 I will discuss an initial distinction made by Douglas in his seminal article on moral bioenhancement. Douglas explores possible objections to moral bioenhancement on the grounds that it could impact personal identity. However, he avoids providing an account of which conception of identity would be vulnerable to moral bioenhancement, opting to rather approach the matter by distinguishing between strong and weak changes to personal identity. He posits that only the former would pose unacceptable ethical challenges to moral bioenhancement.

The distinctions between different conceptions of identity that have been explored by both Parfit and DeGrazia, deemed relevant to the issue at hand, will be discussed in section 4.2.2. In sections 4.2.3 and 4.2.4 I will then address specific concerns regarding the impact that moral bioenhancement may have on personal identity that have appeared in the literature. These concerns include: the threat to self-conception, the inauthenticity of any potential identity changes, the threat to 'inviolable core characteristics' and the threat to moral identity. In section 4.2.5 I will then discuss an important distinction made by Focquaert and Schermer that also has relevance for the concern for moral autonomy. Focquaert and Schermer argue that interventions that are passive, such as deep brain stimulation (DBS) – which has been suggested as a possible mechanism of moral bioenhancement – pose more of a risk to personal identity than those that are active, such as traditional moral education. In particular, they argue that such interventions would be particularly problematic if they resulted in hidden changes to personal identity. This matter has particular relevance for the concern for moral autonomy and will therefore be discussed in detail. In section 4.2.6 I will then conclude with further discussion of DBS, in order to investigate how such an intervention, currently used to treat specific neurological conditions, may impact personal identity.

4.2 The concern for personal identity

The quest to pinpoint what qualities, capacities or components constitute personal identity is an ongoing and vast subfield within philosophy, with relatively scant consensus having been achieved. There are a variety of competing contenders and interpretations and this is reflected in the limited discussions of personal identity within the moral bioenhancement literature[36]. In general, throughout their lives, individuals are subject to events that impact upon, and, to a certain extent, alter their identity. This may occur rather rapidly, due to a traumatic event, or over a protracted length of time where an individual perceives herself to have become a different person to who she once was. However, despite 'feeling' different, there is nevertheless a continuation of something that is sufficient for her to know that she is the same person, and for others to similarly identify her as the same person. Even in cases where a major or instantaneous transformation of personality has taken place, it is very seldom the case that one would posit that this has produced an entirely different person. Generally, it is only in cases of neurological damage or disease that this occurs and even then, the notion of 'entirely different' may be used in a metaphorical manner.

Despite there being little consensus regarding what, precisely, constitutes identity, there seems to be agreement that one's personal identity is inextricably linked with who one is – one's selfhood – and in this regard, it seems to have intrinsic personal value. In other words, if one's personal identity is conceived as one's selfhood, individuals would generally have a decided personal interest in the continuation, and protection, of their identity from external, *unwanted* changes. Thus, an intervention that fundamentally alters an individual's identity, in a way that is unanticipated, could be an affront on their selfhood. By this, I am not arguing that individuals do not have aspects of their personal identities that are negative, nor that many individuals do not desire to change such problematic aspects. I am also not claiming that identities are, or should be, fixed or enduring in nature. If I happen to problematise my levels of empathy, for example, and opt to undergo an intervention in order to increase these levels, this may produce changes to my personal identity. I may come to perceive myself as a caring person for whom the welfare of others is of fundamental importance and this may produce changes in my conduct which in turn affect my self-conception. Furthermore, I would presumably welcome these changes as my having undergone such an intervention would have been motivated by a desire for such a change.

[36] In their comprehensive literature review, Specker et al identify the potential threat to personal identity as one of the neglected areas in the moral bioenhancement literature that warrants further attention (2014:14).

The concern for personal identity, however, is not directed at cases such as this that produce expected, and thus, positive results. Rather, in the context of moral bioenhancement, the concern is that the interventions utilised to morally bioenhance may produce unforeseen effects on personal identity, in the sense of negatively impacting upon "self-understanding, well-being, and social and familial relationships" (Specker et al., 2014:14). In cases of the treatment of a debilitating neurological or psychological condition where an individual is subject to suffering, the moral justification for such procedures and their potential side effects is a more straightforward matter. Any negative side-effects of such procedures, such as the potential impact upon personal identity, would have to be balanced against the benefits associated with the alleviation of suffering that they produce. However, when these procedures are proposed as a means of improving, rather than 'correcting', psychological functioning, or in the case of moral bioenhancement, improving moral functioning, the matter becomes more ethically complex.

Threats to personal identity are a prevalent focus in the area of neuroethics, and, more specifically, they feature in discussions related to the fields of psychopharmacology and neuroscience. Thus, we can turn to some of the discussions in these areas to shed light on the problem of personal identity and moral bioenhancement. Some of the areas that are investigated here are the ways in which the manipulation of neurotransmitters to regulate pathological emotional affect may affect personal identity, as well as the effects of more invasive procedures such as deep brain stimulation (DBS)[37]. A discussion of the potential impact upon personal identity of DBS is warranted due to the fact that this is one of the potential mechanisms of moral bioenhancement that has been mentioned by its proponents (DeGrazia, 2014:362; Persson & Savulescu, 2013:125). However, while both DBS and psychopharmacological interventions are discussed as future potential mechanisms of moral bioenhancement in the literature, their potential effects on personal identity are addressed in a rather perfunctory manner. This matter will therefore be discussed in more detail below. Furthermore, in order to investigate whether or not moral bioenhancement could affect personal identity, it is also necessary to examine the particular conception of identity that is at stake,

[37] DBS has been described as akin to a pacemaker for the brain, as it involves the implantation of an electrode which discharges an electrical current in order to alter deviant brain activity. It is generally used to alleviate the symptoms of conditions such as Parkinson's disease, and other movement disorders, but it has also had success in treating debilitating and chronic affective disorders such as major depressive disorder and obsessive-compulsive disorder (Deep Brain Stimulation for Movement Disorders, 2016). Studies indicate that deep-brain stimulation of the amygdala has an effect on lowering levels of aggression in individuals (Franzini, Marras, Ferroli et al, 2005:83), which is why it has been considered as a potential mechanism of moral bioenhancement. Thinkers, such as Douglas, who regard the goal of moral bioenhancement to be the amelioration of counter-moral emotions, would regard the reduction of aggression as a form of moral enhancement.

as the notion is itself notoriously complex and, as mentioned above, it may be interpreted in a variety of ways.

4.2.1 Strong versus weak identity changes

At the inception of the moral bioenhancement debate, in his seminal article on the subject, Douglas addressed how moral bioenhancement could potentially affect identity. He draws attention to the fact that identity changes are frequently a concern associated with enhancement in general, and therefore argues that an investigation is warranted regarding how moral bioenhancement – particularly because it would entail brain interventions – could produce impacts upon identity (Douglas, 2008:239). Rather than opting to identify a particular interpretation of identity that may be threatened by moral bioenhancement, Douglas simply distinguishes between strong versus weak identity changes, where the former would indicate that a change results in an entirely, or considerably, different person. Whilst Douglas does not explicitly state it, he implies that if moral bioenhancement produced strong identity changes then this would be a negative and unacceptable consequence (2008:239). However, he posits that it is not self-evident that the type of moral bioenhancement that he is suggesting, namely, the attenuation of counter-moral emotions, would in fact produce such strong identity changes.

What is more likely, Douglas posits, is that moral bioenhancement would produce weak changes in identity where this could entail changes in "some of…[the] most fundamental psychological characteristics – characteristics that are…central to how…[a person] views [herself] and [her] relationships with others, or that pervade [her] personality" (2008:239). Whilst changing identity, even in this weaker manner, would not be regarded as ethically unproblematic, particularly if such changes were unanticipated, Douglas does not perceive it as providing a conclusive enough reason for choosing to not morally bioenhance[38]. This is because he posits that we generally "have reasons to preserve our fundamental psychological characteristics only where those characteristics have some positive value" (Douglas, 2008:239). In other words, the argument is that changing aspects of one's personality, such as attenuating certain counter-moral emotions, might be a good thing if such emotions impact negatively upon one's life. However, as Douglas points out, this may not necessarily be the case; an individual may regard the fact that he experiences counter-moral emotions, such as disproportionate aggression, as unproblematic, or even enjoyable. On the other

[38] The view that weak identity changes would be regarded as relatively ethically unproblematic is shared by a number of other thinkers (Specker et al. 2014, Focquaert and Schermer, 2015a, DeGrazia, 2005, Baylis, 2013; Schechtman, 2009).

hand, if one holds the view that a heightened sense of moral depth and acumen will be beneficial for an individual, then this would support Douglas' claim.

4.2.2. Numerical identity versus narrative identity

Specker et al. argue that whilst there are a variety of conceptions of identity, the type of identity in question in the moral bioenhancement debate is "narrative identity rather than numerical identity" (2014:14). Narrative identity may be defined in various ways, however, in its most general interpretation, it implies the identity or self-conception that arises from the way in which individuals unify, integrate and organise all their past and present life experiences, beliefs and values, as well as their future aspirations, into a coherent account or story that is meaningful to them. MacIntyre discusses a similar interpretation of personal identity as referring to "that identity presupposed by the unity of the character which the unity of a narrative requires" (2007:218). By this he argues that a:

> narrative concept of selfhood requires…on the one hand [that] I am what I may justifiably be taken by others to be in the course of living out a story that runs from my birth to my death; I am the subject of a history that is my own and no one else's, that has its own peculiar meaning…To be the subject of a narrative that runs from one's birth to one's death is…to be accountable for the actions and experiences which compose a narratable life. It is, that is, to be open to being asked to give a certain kind of account of what one did or what happened to one or what one witnessed at any earlier point in one's life than the time at which the question is posed (MacIntyre, 2007:217-218).

Narrative identity may be distinguished from numerical identity. Parfit distinguishes between numerical and qualitative identity in order to explicate an understanding of what we mean by 'sameness'. Numerical identity refers to something that is "one and the same" thing (Parfit, 1995:13). Parfit gives the example of a white billiard ball that may appear identical to a second white billiard ball but is nevertheless not the same ball (ibid.). We would say that the two balls are identical qualitatively but not numerically. If we alter the second ball by painting it black, for example, this ball would then no longer be qualitatively identical to the first white ball, but it will remain numerically identical to its former self. In other words, it is still the same ball, albeit, now a different colour. Parfit then discusses this difference in terms of an individual who has suffered a brain trauma and has experienced irrevocable changes in personality. Those close to the individual would then posit that since the accident, he is an entirely different person, based on the view that his personality or character has changed. However, Parfit argues that what they mean here is that he is qualitatively different, rather than numerically different (1995:14). Qualitative identity, defined by Parfit as referring to "the kind of person one is, or wants to be" (1995:14), is

associated with a psychological account of identity and is therefore similar in kind, but is not identical to narrative identity.

DeGrazia also comments on the above distinctions in an article that precedes the moral bioenhancement debate, and aims to discuss the potential impact on personal identity of enhancement in general (2005). He argues that identity discussions within the enhancement debate tend to conflate two different accounts of identity – numerical identity and narrative identity – which leads to much confusion (2005:264). Analytic philosophy has tended to focus primarily on numerical identity, which DeGrazia defines as "the relationship an entity has to itself over time in being one and the same entity" (2005:254). Such an interpretation is able to help elucidate philosophical puzzles and questions regarding how the existence of something is able to persist despite various transformations. In the context of biological enhancement, accounts of numerical identity therefore focus on what changes, if any, would be so great as to result in an entirely different person in a numerical sense (DeGrazia, 2005:265).

DeGrazia argues that numerical identity may be understood in terms of either a psychological or biological approach. The more popular psychological approach can be interpreted in a variety of ways, but, in its most simple interpretation it associates the maintaining of numerical identity with "some sort of psychological continuity" (DeGrazia, 2005:265). This continuity could refer to having uninterrupted memories of oneself, and of one's former intentions and experiences, or it could be associated with maintaining basic "beliefs, desires, and character traits…[or] psychological capacities" (ibid.), even in the absence of memories of earlier selves. At its most basic level, it could simply entail possessing "the capacity for conscious experience" (ibid.).

The biological approach to numerical identity would argue that one maintains numerical identity as long as one occupies the same body. Thus, while the biological approach to numerical identity would view a patient in a persistent vegetative state as the same person, on a psychological account the patient would potentially no longer be the same person, as their selfhood would have been eradicated. While biological numerical identity would possibly be threatened by radical forms of bioenhancement[39], it is virtually impossible to see how it could be threatened by moral

[39] An example of radical bioenhancement that would threaten biological numerical identity, if it ever became possible, would be the uploading of the human mind. This is a possibility explored by transhumanists and proponents of radical biological enhancement as a means of freeing humanity from the constraints of biological existence. Uploading, simply put, "is the process of transferring an intellect from a biological brain to a computer" (Bostrom, 2003:17). There are various means that are posited as potentially leading to such a possibility; however, most of them ultimately depend upon the creation of superintelligence, and the ability to reverse engineer the human brain. It is posited, by thinkers involved in the field of Artificial Intelligence, that reverse engineering the human brain will provide us with a blueprint

bioenhancement. Moreover, it would also be difficult to argue, on a psychological account of numerical identity, that enhancing an individual's moral dispositions would threaten her numerical identity, thereby creating an entirely new individual (DeGrazia, 2005:267). As DeGrazia points out, an individual would, of course, remember herself before her enhancement; thus, unless consciousness was eradicated, it seems that psychological numerical identity would remain intact.

DeGrazia, however, argues that not only is there confusion between different types of identity, but in identity debates there is also a tendency to "overestimate[e] the strengths of a psychological approach to our numerical identity and underestimate[e] the strengths of a biological approach" (2005:264). DeGrazia supports the biological approach to numerical identity as its materialist conception of the human being as an 'organism' assists us in avoiding conflating the quest to identify what confers human value or essence "with the metaphysical issue of numerical identity" (2005:266). However, despite his views in this regard, DeGrazia posits that the form of identity that is actually at stake in discussions of enhancement is not numerical identity – whether psychological or biological. Rather, what matters to individuals when they refer to a loss of, or impact upon, their identity, is narrative identity. This, he posits, refers to individuals' "self-conception: [their] most central values, implicit autobiography, and identifications with particular people, activities, and roles" (ibid.).

Narrative identity is, of course, seemingly strongly related to a psychological account of numerical identity. However, DeGrazia argues that the two types of identity are nevertheless distinct as they have different foci and aim to answer different questions. Numerical identity attempts to establish conditions for sameness in the same organism or entity in the face of changes. In other words, it aims to decipher what conditions would have to be met for an organism or entity to become an entirely different one. In this regard, DeGrazia identifies numerical identity as dealing with "metaphysical or conceptual" questions (2005:266). Narrative identity, on the other hand, attempts to identify "what is most central and salient in a given person's self-conception" (DeGrazia, 2005:266) and in this regard, it is "value-laden and *inherently* psychological" in nature (ibid.). In other words, it is an internal state that is based upon interpretations and beliefs – and evaluations

for intelligence and enable us to replicate its mechanisms. Once the mechanisms of the brain are fully comprehended, the next step will be downloading the brain, which would entail "scan[ning the] brain to map the locations, interconnections and contents of all the…neural components and levels. Its entire organisation…[could] then be re-created on a neural computer of sufficient capacity, including the contents of its memory" (Kurzweil, 2001:26). At the point at which we are able to scan the brain at the minute level required in order to fully understand its workings, the intricacies of the process of downloading or replicating it mechanically will supposedly be solvable, according to transhumanists.

thereof – regarding what is salient for a given individual about herself, how she perceives herself to be and how she thinks others perceive her.

4.2.3 Narrative identity and self-conception

It seems obvious that if narrative identity is strongly associated with an individual's self-conception over time, then a discussion of how moral bioenhancement could impact upon personal identity would have to examine how it could affect the qualities that inform self-conception. This would include investigating whether or not moral bioenhancement would affect how individuals view themselves, and, whether or not it would impinge on those qualities that they perceive as constituting their selfhood. DeGrazia discusses an example of how the enhancement of physical capabilities or physical beauty could deeply impact a person's self-conception. However, he also draws attention to the fact that this would be true of any acute change, such as losing the use of a limb or suffering a stroke, for example. This observation aside, however, he examines two specific objections to enhancement that are strongly associated with a concern for identity; namely, the argument that it is inauthentic and the concern that it will threaten "inviolable core characteristics" (DeGrazia, 2005:269). Arguments that view enhancement as representing a threat to authenticity generally associate the latter with the ideal of "self-creation" through effort or "being true to oneself" (DeGrazia, 2005:268). In terms of authenticity as self-creation however, freely utilising enhancement technologies to change oneself in some way could be interpreted as a form of self-creation, rather than a threat to it.

On the other hand, if authenticity rather refers to an acceptance of who one is – or being 'true to oneself' – including one's identity, it would appear that using enhancement could result in distinct changes in this regard, which according to this interpretation would be inauthentic[40]. As DeGrazia points out, however, one of the problems with such arguments is that they are characterised by confusion between numerical and narrative identity (2005:269). Altering someone's numerical identity clearly represents a strong identity change, to use Douglas' distinction, as it would essentially destroy the person. The consensus is that this would be a clearly negative, and ethically untenable, outcome. However, it is clearly not numerical identity that would be at stake in cases

[40] This type of criticism is frequently lodged by those with a bioconservative agenda. Those who hold a bioconservative view are generally opposed to biological enhancement in general, as well as other biomedical interventions, such as stem cell research which destroys human embryos in the process; the destruction of embryos in general, including abortion, as well as practices such as euthanasia. Bioconservatives do support the use of genetic technologies for the treatment of inheritable genetic disorders, however, they oppose any attempts to enhance human capacities due to their view that the human being in its present state should be left as is. According to bioconservatives, attempting to alter or 'improve' our genetic structure amounts to hubris, playing God or tampering with nature, and risks destroying our humanity (Kass, 2002; Sandel, 2007).

of enhancement, as the individual would remain the same person in a numerical sense. What is rather at stake is narrative identity, the individual's self-conception over time, which could be affected by enhancement. However, it is accepted as relatively unproblematic that our narrative identities are continually transformed in diverse ways due to a variety of influences. This is particularly true of changes that result from freely made choices, such as the decision to enhance oneself in some way. Thus, an argument that opposes freely chosen enhancement on the basis that it may impact narrative identity would have to indicate how the supposed inauthenticity it produces is different in kind to the impact of general changes to identity experienced in everyday life, particularly those changes that occur as a result of factors over which we have no control.

The second criticism, that bioenhancement will in some way threaten "inviolable core characteristics" (DeGrazia, 2005:269), is also a common bioconservative argument. Such arguments are deeply problematic, however, due to the unsubstantiated assumptions that they make. One such problem with this line of criticism is, how do we identify which of these supposedly core characteristics are inviolable and thereby should not be altered in any way; and why are they inviolable? Furthermore, if an individual has made an autonomous, informed choice to alter what is viewed as a core characteristic, upon what basis could this be regarded as inauthentic? In these kinds of 'arguments', identity, or the notion of "core characteristics" (ibid.), may serve as a placeholder for a variety of inchoate notions such as human essence, human nature or a metaphysical 'true self', the nature of which must simply be intuited. Regardless of whether one supports the existence of the afore-mentioned notions, their supposed inviolability would nevertheless be constitutive of numerical identity, rather than narrative identity; which is the type of identity that would be more likely to be affected by bioenhancement. Furthermore, those who hold the view that it is morally objectionable to alter these supposedly core characteristics must provide more substantial arguments as to why this is so, rather than offering unsubstantiated moral proclamations that, in the absence of supporting arguments, amount to distinctly circular claims.

4.2.4 Narrative identity and moral identity

While, as mentioned above, there seems to be nothing intrinsically wrong with altering aspects of our narrative identity, Faust makes an interesting observation that requires further investigation. She posits that moral identity is constituent of narrative identity and that we care about the former "as it relates to our moral integrity – we choose to act in ways that are consistent with our moral beliefs in order to maintain our moral integrity" (Faust, 2008:403). Jotterand makes a similar point and alludes to the concern that moral bioenhancement is ethically problematic because it may

threaten "moral identities" (2014:2). As mentioned in section 2.3.5, he distinguishes between the possession of "character traits…and having character" (Jotterand, 2011:8), where the former refers to a variety of behaviours that may, or may not, possess moral relevance. I may have a particular character trait, such as conscientiousness, but only employ it, for example, in my workplace rather than in my attitude towards my personal relationships. Thus, simply possessing that trait will not necessarily signify that I employ it in a morally relevant manner or context. For Jotterand:

> having character implies having a moral identity; it implies a person's moral strength to establish a set of behaviours deemed adequate in projected circumstances. It qualifies one's moral agency and presupposes one's capacity of self-determination. Agency (reasons, motives, intentions) and action constitute the two elements that refer to having character (2011:8).

In other words, particular character traits may be necessary for possessing character but they are not sufficient. Having character includes not only possessing the morally relevant character traits, but also the notions mentioned above; namely: "a moral identity…moral strength…[and] moral agency…[that inform or lead to] action" (ibid.). He posits that it is character traits and behaviour that are the target of moral bioenhancement, but it is having character that is, arguably, the basis of morality itself. In other words, what the above thinkers are implying is that if our moral identity or integrity is associated with the coherence between our moral beliefs and our actions, then an intervention that produces behavioural changes without concomitant changes in our moral beliefs could be problematic. Moral identity, so interpreted, presents us with a construal of identity that may be threatened by moral bioenhancement, and which is inextricably linked with the Harris concern for moral autonomy, where a requirement of the latter would be possessing adequate reasons and beliefs that can explain our behaviour. Thus, where one's personal identity or self-conception is informed by one's considered conception of what is true and good, any changes that would result in an identity that is incongruent with this previously considered conception could be ethically problematic, and, as I will argue in chapter 5 could also threaten personal or moral autonomy.

4.2.5 Potential impacts on narrative identity: hidden changes

Returning to Specker et al.'s discussion of identity, they provide a decidedly psychological interpretation of narrative identity as referring to "an individual's most central and salient characteristics (e.g., motivations, beliefs, values, desires, character traits) that together comprise their self, and needs to be understood within the dynamics of psychological change" (2014:14). Of course, as Specker et al. point out, our narrative identity, so construed, is not static throughout our lifetime; it is continually transformed and impinged upon as we react to particular occurrences and

interact with other individuals (Specker et al. 2014:14). These changes are the basis of the kinds of observations that support the view that someone is now not the same person as they once were. However, what matters to individuals is that they are able to assimilate any life changes into their self-conception or identity – their life story – in an intelligible and meaningful way (ibid.). One such way that this would be possible is to be able to link any personality changes with specific events. In other words, an intelligible incorporation of a change to an individual's identity would entail being able to provide reasons to explain the source of such a change.

Most individuals are accustomed to experiencing changes to their personal identity and these are generally assimilated without major psychological upheaval. As mentioned above, and in agreement with Douglas, a number of thinkers have argued that if moral bioenhancement, and bioenhancement in general, were to produce such smaller or weaker changes in identity, then, if this occurred in a similar manner to the kind of identity changes that individuals are confronted with in the normal course of events, this would be ethically unproblematic (Specker et al. 2014, Focquaert and Schermer, 2015a, DeGrazia, 2005, Baylis, 2013; Schechtman, 2009). However, as pointed out by Specker et al., because moral bioenhancement would invariably impact upon core "moral dispositions or behaviour" (2014:14), this may result in unpredictable, and possibly major disruptions to narrative identity which would elicit acute ethical concern.

Focquaert and Schermer discuss identity concerns in some depth and also identify narrative identity as the particular conception of identity that is vulnerable to potential impact from moral bioenhancement. They utilise Schechtman's interpretation of narrative identity, understood as referring to a person's "actions, experiences, beliefs, values, desires and character traits" (in Focquaert & Schermer, 2015a:146). Furthermore, they share the above-mentioned view that our narrative identity is subject to constant change and that minor changes are generally assimilated into one's narrative in a relatively unproblematic manner, and in accordance with maintaining "one's sense of self" (ibid.). Major changes that severely interrupt or are at odds with narrative identity *could* pose ethical hurdles. The extent to which moral bioenhancement could do this would vary according to the nature of the intervention; it is a particular type of intervention that would be more likely to produce severe impacts on identity. Here, they distinguish between, active and passive moral enhancements and argue that this distinction is ethically meaningful (Focquaert & Schermer, 2015a:145).

Generally, enhancement is framed in terms of a distinction between direct and indirect enhancement, where the latter is associated with traditional moral enhancement via education and socialisation and the former with biological enhancement. Focquaert and Schermer argue, however, that this distinction can be reformulated in terms of a deeper, and more relevant, foundational distinction between interventions that are active versus those that are passive. Traditional enhancement is 'active' because it is a process that requires individual effort and ongoing involvement. However, it could also be described as indirect due to the fact that it results in changes in attitude which gradually affect mental states rather than acting directly, and immediately, on the brain (Focquaert & Schermer, 2015a:144)[41]. Passive moral enhancement would include moral bioenhancement interventions which require no, or at least minimal, individual contribution or participation as they act directly on the brain (ibid.). With this distinction in mind, Focquaert and Schermer examine potential interventions that could impact morality, as occurring on a spectrum that moves from active to passive mechanisms. We can then visualise this spectrum, where traditional moral education lies on the extreme left, moving to talk-based interventions, such as cognitive behavioural therapy, through to the administering of psycho-pharmaceuticals, and then, finally moving to interventions such as deep brain stimulation on the extreme right. Focquaert and Schermer argue that it is the interventions that are passive on the far right of the spectrum, that pose the greatest threat to narrative identity. However, they posit that even if such interventions do produce acute, and disruptive, identity changes, it is not self-evident that this would imply that such an outcome is unethical (Focquaert & Schermer, 2015a:147). Through various supportive mechanisms, such abrupt identity changes could be assimilated. The type of narrative identity change that would pose a distinct ethical obstacle would be what Focquaert and Schermer describe as hidden changes to narrative identity (ibid.).

Such hidden changes refer to smaller changes that are either not recognised by the individual in question, or, are perceived as being more minor by the individual than they are by those who are close to them (Focquaert & Schermer, 2015a:147). Examples of such occurrences are drawn from patients who have undergone DBS which has produced subtle, long term changes that have deeply affected their personal relationships. What was interesting in one particular case was that the changes in personality were not recognised by the patient in question when the brain implant was

[41] However, in a later paper, Focquaert and Schermer explicitly point out that they "do not equate direct interventions with passive ones and indirect interventions with active ones" (2015c). While they posit that "direct interventions are more likely to be passive and therefore more likely to be problematic" (ibid), this is not always the case. In fact, due to various problems associated with the direct/indirect distinction, they argue that the distinction between active and passive interventions is more useful as a means of explicating the ethical status of interventions and should supplant the former distinction.

switched on. Only when it was switched off, and his Parkinson's symptoms returned, did he concede that he was exhibiting problematic behaviour. However, even subtle changes, such as increased "irritability, impatience" (Focquaert & Schermer, 2015a:147) and concentration difficulties may have an erosive effect on relationships, particularly when not acknowledged by the patient.

As Focquaert and Schermer point out, narrative identity may be evaluated from both the perspective of the person in question and from the perspective of others. The individual evaluates her own narrative by whether or not she is able to assimilate any perceived changes and anomalous feelings or behaviours. The primary way in which she is able to do this is by being able to provide explanations for any changes. However, this presupposes a certain amount of self-insight and it is possible, as Schechtman notes, that there can be "a dissociation between one's implicit narrative self and one's explicit narrative self" akin to a type of "self-blindness" (in Focquaert & Schermer, 2015a:148). In the absence of actually noticing any subtle personality changes, the individual wouldn't perceive any need to assimilate them into her narrative. In other words, the individual would perceive no need to deliberate upon any such changes, and in this way, such a change would bypass her faculties of reasoning. Focquaert and Schermer see passive interventions as the most likely source of such hidden changes.

As mentioned above, such "self-blindness" (Focquaert & Schermer, 2015a:148) would be problematic not only in terms of the negative impact it may have on relationships, but, more importantly, it could be interpreted as a form of "inauthenticity that threatens the autonomy of the self" (ibid.). Focquaert and Schermer concur with DeGrazia that to avoid identity inauthenticity and preserve autonomy, it is imperative that an individual be able discern which changes to assimilate and which to reject. Being able to do this presupposes that the individual firstly can recognise the changes and relate to them, and, secondly, that she can pinpoint the source of the changes. It is clear that hidden changes will not meet these criteria for authenticity. However, it is difficult to see how interventions that produced such effects would ever be considered as viable candidates for moral bioenhancement.

Caouette has responded to the threat to narrative identity, and thus authenticity, raised by hidden changes that Focquaert and Schermer discuss. In particular, he looks at their claim that direct interventions could bypass individuals' rational and reflective capacities which would enable them to adjudicate such changes (Caouette, 2015). Caouette, however, is of the view that Focquaert and

Schermer have overstated this threat to authenticity. Regarding explicit personality changes, he argues firstly that we often find ourselves experiencing inexplicable emotions or moods with no apparent origin or cause. However, we always have a choice regarding how we will respond to such affective states. In the same way, if we experienced such changes in emotion or disposition due to a direct or passive intervention, it would not be the case that the way in which we respond to this would be predetermined. It does not follow that the specific nature of the source of such a change determines whether or not our autonomy regarding how we respond is eradicated.

Regarding Focquaert and Schermer's requirement that authenticity presupposes the ability to coherently explain such changes, Caouette argues that there is no reason why this ability would be threatened by a passive intervention. For example, we are often unclear as to why our feelings regarding a certain matter, or person, change, however, we are nevertheless, generally able to assimilate these changes into our narrative identities. What matters, he argues, is keeping intact the ability to choose to either act on these changes or not (Caouette, 2015)[42]. Furthermore, he posits that even in cases where an intervention such as DBS is not voluntary, such as would be the case if it was utilised to treat psychopathologies in criminals, it would not rule out the individual being able to incorporate any changes in moral dispositions, for example, into his narrative. He equates such involuntary changes with the changes in perspective that arise in one's life due to external factors over which one has little control. Some examples of events that could result in identity changes would be most traumatic experiences such as the loss of a loved one, divorce, loss of one's employment, or the particular example that Caouette mentions of being bullied as a child. After experiencing abrupt changes to one's self-conception as a result of such an experience, one will presumably remember who one was before and utilise one's 'new' moral dispositions to make sense of any new feelings and possibly welcome them.

What is of particular interest in their discussion, however, is the way in which Focquaert and Schermer's argument clearly illustrates the close relationship between personal identity and autonomy. In this regard, their distinction is loosely aligned with the Harris line argumentation,

[42] Focquaert and Schermer have responded to Caouette's claims by arguing that he has not correctly understood the risk to identity, and thus to authenticity, posed by radical changes to identity (2015b). They point out that in the case of slight, or isolated, identity changes, Caouette is correct in his claims that an individual may be able to assimilate such changes into her identity and choose whether or not to act upon them. However, in the case of major identity changes, "if 'we' ourselves have changed considerably, it is not clear anymore that our choices will really be authentically ours" (Focquaert & Schermer, 2015b). In other words, if an intervention drastically alters the "values, desires, propensities or…outlook" (ibid) of an individual, then one could argue that the person evaluating such changes and deciding whether or not to assimilate them and act upon them may no longer be the same person. This would be even more pertinent in the case of hidden identity changes.

which is the claim that an intervention threatens moral autonomy – in their case narrative identity – if the changes it produces have no cognitive content. Focquaert and Schermer discuss how an active, non-invasive intervention such as cognitive behavioural therapy may also result in major behavioural changes and thus changes to identity. However, while the impact of such identity changes may be equally comparable in magnitude to passive, or direct, interventions, such as DBS, the difference between the two is that the effects of active interventions occur gradually and can therefore be reflected upon at length and either assimilated or discarded by the individual. Direct, or passive, interventions that produce major and instantaneous disruptions to identity, on the other hand, leave little opportunity for gradual revision and cognitive consideration. Furthermore, if such interventions were not reversible, this would be even more problematic. In this way, there is congruence between the concern that moral bioenhancement could threaten something fundamental and intrinsically valuable, namely, narrative identity – understood as one's self-conception or moral identity – and moral autonomy. For Focquaert and Schermer, their concern for identity is that passive interventions, proposed as mechanisms of moral bioenhancement, could produce character changes and disruptions that could be enduring because they would be seemingly inaccessible to cognitive consideration. Their concern for identity is therefore, in essence, a concern for autonomy where the latter is strongly associated with the ability to self-determine. If their argument is correct, the question that needs to be investigated is whether or not moral bioenhancement could produce the kinds of irrevocable disruptions to narrative identity that they fear. Answering this question, and resolving the matter, would require empirical investigation.

4.2.6 Narrative Identity and deep brain stimulation

A number of thinkers have explored the potential threat to identity posed specifically by deep brain stimulation (DBS). As mentioned above, DBS is used to treat Parkinson's disease and other movement disorders and has also been suggested as a potential mechanism of moral bioenhancement (DeGrazia, 2014:362; Persson & Savulescu, 2013:125). However, a commonly noted side-effect of DBS is the impact that it has on personality and/or identity (Schechtman, 2010). As Schechtman points out, in cases where DBS is used to treat mood disorders, such as major depressive disorder, personality changes may be the explicit goal of such an intervention, however, more often, they are undesirable and "unintended side-effect[s]" (2010:133). Such changes can occur rapidly as well as in a more gradual manner and may cause great personal distress for patients undergoing DBS.

Schechtman discusses some examples of acute changes in demeanour exhibited by patients after DBS. She reports that patients have described changes in emotional response as sudden as if a switch had been flipped (2010:135). One such patient had been experiencing long-term depression which completely disappeared, with such rapidity, that it elicited intense anxiety, despite the positive nature of the change. The fact that such positive changes elicited anxiety is indicative that some fundamental challenge to the patient's self-conception or identity could have been the source of his unease. This is understandable, as an individual suffering from chronic depression would presumably have been required to assimilate his condition into his identity, and thus, any rapid changes in this regard could be experienced as a threat to his selfhood or identity. Other patients describe how feelings can change instantly from happiness and laughter to sadness, and vice versa, during the testing and placement of the deep brain stimulator (Schechtman, 2010:134).

However, it is not only in invasive interventions, such as DBS, that such acute changes have been noted. Schechtman also discusses the ways in which the SSRI, Prozac, has produced similar extreme changes in emotion and personality (2010:134). Schechtman explains how the reaction of unease to such changes is connected with "a perceived threat to conceptions of identity and agency that have been deep and long-standing parts of Western culture" (ibid.). These kinds of instant changes are disturbing because they undermine the perception of the extent of control that we have over ourselves and thus erode the view of the existence of a true or authentic self that we believe – correctly, or not – is the product of our own volitions. Furthermore, changes of this nature also undeniably confirm the immense role played by physiological processes on "personality, intellectual performance, and social success – that heretofore we as a society have resisted" (ibid.). As Schechtman points out, the fact "that one's very psychological identity can be altered by chemical means raises questions regarding how it is then possible to think of oneself as an autonomous, self-directed being" (2010:135) and this, in turn, elicits great unease. In other words, in the majority of individuals, who we can presume have not been exposed to philosophical discussions of freewill and determinism, there is a common perception that we are freer from causal influences than we, in fact, are, and when this perception is challenged, it may be deeply unsettling.

Schechtman describes the various ways in which identity may be conceived and concurs with the prevalent view that, in terms of the potential threat to identity envisaged by DBS, it is narrative identity that risks being impacted upon. She defines identity as not referring to stable individual characteristics or personality traits but rather our "selfhood…[as] tied…[to] our ability to understand ourselves and others in narrative terms. We are selves – and construct identities –

insofar as we experience and live our lives as narratives" (Schechtman, 2010:137). The construction of these narratives is not performed with explicit awareness but rather refers to the fact that we interpret ourselves as subjects who exist in time and space with personal and historical contexts that all play a role in who we are, and who we will be. This ongoing story of our lives is, of course, impinged upon by random, and sometimes haphazard, events.

Schechtman argues that in comparison, if we take an inanimate object, it is generally possible to acquire all the relevant facts regarding its history. Namely, we can identify how the object came to exist and specifically trace and account for any changes in its form. We cannot do this with human beings as their narratives are informed by a dynamic complexity that is an amalgamation of "purposes; goals and plans…emotions, beliefs, values, and desires that develop and change in response to circumstances and constrain action; and complex relationships with others" (Schechtman, 2010:137). When an individual experiences a change in her emotions or dispositions, this may halt or impede the "narrative flow" (ibid.) of her identity, thus requiring that she assimilate this interruption in a coherent manner within her narrative. This assimilation is achieved through seeking understanding and acquiring explanations for any such changes. In the case of DBS, when this change occurs in a jarring and instantaneous manner, the only explanation that is available to the individual as a means of assimilating this interruption into her narrative is the kind of "mechanical explanation" (ibid.) that is associated with inanimate objects. This may contribute towards the feelings of alienation that are described by patients after receiving DBS.

However, in contradistinction to static, psychological accounts of identity, narrative identity is flexible and adapts and responds in a continual interplay between the individual and environment. Thus, as Schechtman points out, one's identity may persist even in the face of major disruptions, particularly if one is able to interpret such changes as "self-expressive and self-directed" (2010:138). What she means by this, is that whilst the most immediate or proximal cause of any changes may be identified as a mechanical one; namely the DBS one has received, if one takes a wider perspective and moves further back, the change may also be attributed to a distal cause, namely, the willingness and proactivity to address whatever condition, or state, caused the individual to seek treatment, or enhancement, in the first place.

Baylis also explores the threat to identity posed by DBS, and correctly points out that whether DBS does, in fact, threaten identity, depends entirely on the particular conception of identity that one has (2013:520). He interprets identity as relational and narrative in character (Baylis, 2013:513).

In contradistinction to psychological accounts of identity that associate it with stable "core inclinations or character traits" (Baylis, 2013:516) and biological accounts of identity that link it with occupying a particular corporeal form, Baylis regards identity as associated with "an individual's lived experience as integrated into her autobiographical narrative" (ibid.). By this he means that an individual's identity is coherent if she is able to give plausible explanations or possess an interpretation of her "history…life situation…and…motivations" (Baylis, 2013:517) that is, more or less, a reflection of reality. He defines relational identity, more fully, as "a dynamic, socially, culturally, politically and historically situated communicative activity (based in narrative and performance) that is informed by the interests, perspectives, and creative intentions of close and distant others" (Baylis, 2013:17). This conception of identity is dynamic in that it is constituted by interplay between an individual's self-conception and the way in which they are perceived by others.

Baylis argues that while identity and personality are often conflated, the two are not the same (2013:516). In distinction to relational or narrative accounts of identity, psychological and personality accounts of identity tend to be more static, and it is this latter –and outdated, Baylis argues – conception of identity that risks being impacted upon by DBS. However, if one interprets identity in such static terms, then one must include, as threatening to identity, any life event that produces dramatic change to one's identity, such as those traumatic events mentioned above, namely, divorce, death of a loved one or loss of employment, to name but a few examples. If one takes Baylis' two criteria that a coherent identity requires that changes can be explained and that they bear a relationship to reality, then there is no reason to believe that DBS couldn't fulfil these requirements. Presumably, a patient experiencing acknowledged personality changes as a result of DBS would be able to provide a plausible explanation regarding the source of any changes because his having consented to the intervention would be an event that would be assimilated into his personal narrative. If one conceives of identity as something dynamic and subject to continual external and internal influences and changes that unfold throughout an individual's life, an intervention such as DBS is merely one additional way in which identity is formed. However, if one associates identity with static personality traits or the existence of a metaphysical, or given, 'true self', then the former may be threatened by such interventions.

Baylis does, however, very briefly, discuss a similar threat to that implicitly identified by Focquaert and Schermer; namely the threat to autonomy or agency that could be posed by an intervention such as DBS. If DBS produced unwanted, or hidden, changes in behaviour and action which

resulted directly from the intervention, rather than the formation of "intentions or beliefs" (Baylis, 2013:524), then this could be grounds for an argument that such an intervention compromises agency. One such example of the way in which this does occur, is a heightened risk for the development of addictive gambling behaviours in patients treated with DBS for Parkinson's disease (Lu et al., 2006; Smeding et al., 2007; Hälbig, 2009). If an individual, post-DBS, engages in such behaviours, not as a result of having chosen to do so, but due to the intervention itself, then it would be difficult to interpret this as anything other than a subversion of his agency.

In terms of its therapeutic usage, an invasive intervention such as DBS, used to treat debilitating symptoms such as those produced by Parkinson's disease, could therefore produce side-effects that must be balanced against the advantages of alleviating such symptoms. However, the fact that DBS is a relatively common form of treatment for Parkinson's disease, despite such agency subverting side-effects, is evidence that it is viewed as ethically permissible to make such trade-offs. Furthermore, utilising DBS to treat severely debilitating mood disorders, where other forms of treatment have failed, would also be ethically permissible, even in the face of the afore-mentioned side-effects. However, it seems highly doubtful that individuals would utilise such interventions, in the absence of coercion or sufficient incentives, to enhance particular moral dispositions deemed in need of strengthening; even if it could be established that such interventions did not impact upon identity.

In this regard, Riis et al. have conducted empirical research that has implications for the likelihood that people would willingly opt to morally bioenhance themselves (2008). The research specifically investigates the likelihood of individuals freely choosing to take pharmaceuticals that will "enhance their own social, emotional, and cognitive traits" (2008:495). Their findings indicate that, particularly in Western culture, there is an enduring belief that certain traits are closely associated with, and inform, a "fundamental, essential self" (Riis et al, 2008:495) or identity. Thus, whether or not individuals would be willing to enhance themselves would seem to depend upon their beliefs regarding their "fundamental selves" (2008:496) or their identities. In particular, emotional dispositions are viewed as constitutive of identity; thus, we could expect reluctance to alter such dispositions, unless, as Riis et al. point out, the enhancement of such dispositions was framed, and marketed, as "enablers of one's true self" (2008:505).

4.3 Concluding remarks

This first part of chapter 4 addressed an important argument against moral bioenhancement, the argument that moral bioenhancement is wrong, in principle, due to the possibly disruptive effects that it could have on personal identity. However, an investigation of the literature revealed that there are numerous conceptions of personal identity. Thus, this concern requires conceptual clarification in order to ascertain which interpretation of personal identity is, in fact, at stake. Secondly, it requires examining empirical evidence from areas in which some of the interventions proposed by the proponents of moral bioenhancement are currently used for therapeutic purposes, as different interventions could compromise identity to different degrees.

While identity may be defined in various ways, it seems that it is generally associated with selfhood in some way. Thus, this leads to the perception that individuals have a vested interest in protecting their identity, their *self*, from strong, unwanted changes. On this account, if moral bioenhancement were to produce strong identity changes, to the extent that this would threaten self-conception, this would be cause for ethical concern. Whether moral bioenhancement would result in such changes, however, would depend upon the nature of the intervention. However, in cases of weaker identity changes, a strong argument may be made that altering negative aspects of an individual's identity, such as certain counter-moral emotions, could be a positive outcome for individuals. If the presence of such a disposition is rejected by an individual and produces inner conflict or has a negative impact on her life, such as compromising her ability to lead her life in a manner that is congruent with what she takes to be true and good, she would generally not regard such a disposition as part of her identity. In such cases, identity changes could be regarded as increasing the agency or autonomy of individuals. This is an argument that requires further substantiation and it will thus be developed further in chapter 5b.

The contribution that a philosophical investigation may make in explicating the concern for personal identity posed by moral bioenhancement would be to provide conceptual clarification regarding the notion of personal identity that is at stake. Here, in the most general sense, it is qualitative identity that is seemingly at stake. Qualitative identity is both descriptive and aspirational in character as it is informed by how individuals conceive themselves – the self that they believe themselves to be – as well as the selves they aspire to be. In particular, there seems to be considerable consensus in the moral bioenhancement literature that narrative identity is the particular type of qualitative identity that would be vulnerable to impacts from moral bioenhancement. This account of identity may also be defined in various ways. It includes not

only an individual's self-conception, but also his interpretations of the ongoing events of his life and his relationship with others. It is therefore not static as it is subject to impacts from continual influences and alterations due to interactions with others and with the world in general.

However, I would argue that whether or not personal identity is interpreted in narrative terms, the core component of any conception of personal identity is the notion of selfhood. Furthermore, I would argue that selfhood is deeply informed by moral attitudinal states, referring to what an individual values and identifies with, as well her conception of how consistent she is in acting in accordance with her moral beliefs. This moral component of identity is more cognitive in nature. In other words, acting in accordance with one's moral identity would require that we act upon specific intentions that we have formed, that are consistent with our beliefs, and, that we are therefore able to provide reasons or justifications for our actions. We may not always do this in a conscious manner, but if pressed, components of our moral identities are those beliefs that we *could* provide explanations for and would be willing to defend if necessary. Furthermore, I would argue that this moral aspect of one's identity, while not fixed, is more enduring in nature. It is closely linked with that part of an individual that enables others to recognise her as a particular individual through time and to observe, in cases of major change, that she has indeed changed.

Regarding changes in identity, it is not change as such that is problematic, but rather, whether or not individuals are able to assimilate any changes into their self-conception and their personal narrative in a way that is acceptable to them. When changes cannot be linked to events in one's narrative or to any pre-existing beliefs, preferences or attitudinal states, then the source of such changes would warrant investigation. However, it is most likely that only a specific type of intervention would warrant caution in this regard. Any interventions that act directly upon physiology or brain states, and therefore do not require any individual involvement to ensure their efficacy, would be potentially problematic. Furthermore, any interventions that produce hidden identity changes would require our utmost attention. This is because hidden changes could produce impacts upon autonomy, where the latter is associated with one's ability to self-determine in a manner that is free from *undue* internal and external constraints[43]. Their unacceptability would not only be due to the fact that they could challenge our autonomy as such, but also due to the impossibility of addressing hidden, and thus unrecognised, changes, other than relying on the reports and interpretations of others. In particular, hidden changes in identity could be at odds with

[43] Of course, on a deterministic account of freedom – an account that I believe to be correct – our ability to self-determine is not absolute, we are subject to causal constraints. I will discuss this matter further in the second part of this chapter in section 4.10.

the requirements of true informed consent, in terms of the fact that consent could not be withdrawn, unless the intervention was reversible and the individual in question heeded the accounts of others and sought a reversal.

Chapter 4b – In-principle objections to moral bioenhancement: the concern for moral autonomy

4.4 Introduction and overview

The concern for moral autonomy forms part of a set of freedom-related concerns that are prevalently discussed within the moral bioenhancement debate. For the purposes of this chapter, which will focus on how the concern is discussed in the literature, we can loosely distinguish between freedom concerns and autonomy concerns[44]. On the one hand, as discussed in section 3.3.1 of chapter 3, freedom concerns are often directed towards compulsory moral bioenhancement on the grounds that it would be an obvious violation of our freedom of choice. Such a concern is, however, not specific to moral bioenhancement. Rather, it typifies any situation in which the freedom of individuals to make reasonable personal, bodily choices is impinged upon by making something compulsory. In the case of moral bioenhancement, thwarting the freedom of individuals – specifically their bodily freedom – to choose to refuse a physiological or neurobiological intervention would be clearly ethically problematic[45].

Critics also argue, however, that moral bioenhancement is wrong on a deeper level, and, in a way that would not be resolved by ensuring that it would only be administered on a voluntary basis. Critics argue that it is wrong, in principle, due to the fact that it will inevitably compromise or eradicate something that holds intrinsic value to humanity, namely, our moral autonomy or agency. As the discussion in this chapter will illustrate, in the debate, terms such as moral autonomy, agency, freedom, liberty, free will, moral responsibility and autonomy are often used interchangeably to emphasise different concerns elicited by moral bioenhancement. Therefore, it

[44] In chapter 5a I will provide a conceptual analysis of the notion of autonomy which will include an overview of traditional conceptions of autonomy as well as an explication of the distinction between freedom and autonomy in section 5.2. My interpretation of this distinction is informed by Berlin's distinction between negative and positive liberty, where the former refers to freedom *from* undue interferences and constraints and the latter refers to freedom *to* do a particular thing or be a particular person, where this is associated with an individual's sense of self-determination (1969). This distinction between freedom and autonomy, in the sense in which it is relevant to the moral bioenhancement problematic, also correlates with the contrast between external and internal freedom. I will, of course, discuss the nature of this distinction in more detail in chapter 5, section 5.2, however, for the purposes of this chapter, it is important to note that a variety of terms are utilised in the literature. Therefore, in this chapter, when I utilise terms such as, freedom, liberty, agency, autonomy and moral autonomy, this is not due to a lack of conceptual precision, but rather, because these are the particular terms utilised by various thinkers in their arguments.

[45] While this is the prevalent view in the literature, the truth of this claim may not be regarded as self-evident. However, in this regard, I am not referring to extraordinary cases in which a medical intervention is performed without the consent of an individual, in order to save his life. Rather, I am referring to a situation in which individuals are forced against their will, either through covert or compulsory administration, to undergo a bioenhancement. Persson and Savulescu's suggestion of putting oxytocin in the drinking water in order to ensure universal coverage of moral bioenhancement would be an example of the latter (2008:174).

is necessary to investigate whether some of the confusion and lack of consensus in the debate may be attributed to the fact that what is at stake is not always explicitly and clearly identified. In this regard, some clarification would be a worthwhile endeavour. However, despite this confusion and ambiguity, one nevertheless has an intuition of what it is that could be eradicated, should the worst fears regarding moral bioenhancement be realised. In this second part of chapter 4, I will address the way in which the concern for moral autonomy has been formulated, discussed and debated in the literature. I will then synthesise these insights, along with relevant findings from chapters 2 and 3, and develop them in chapter 5 to reach a clear conclusion regarding the concern for moral autonomy that is posed by moral bioenhancement.

While there is an abundance of input addressing the concern for moral autonomy, most of the commentary in the literature engages with the arguments of Persson and Savulescu and the responses of Harris. In section 4.5 I will therefore largely focus on the ongoing debate between these authors. In section 4.6 I will re-examine the issue of the role of reasoning versus emotions in morality, particularly in terms of how Harris' claims regarding the dangers associated with emotional modulation have been challenged by insights from the field of neuroscience. However, these latter findings could indirectly support the claims that Harris has made elsewhere; namely, that what is needed is cognitive, rather than moral, bioenhancement.

The concern for the impact of moral bioenhancement on moral autonomy may be addressed in two ways: it may be refuted, or, it can be justified. In other words, one can either construct arguments as to why moral bioenhancement will, in fact, not produce the threats to autonomy that are feared, or, one can provide justifications for any potential impacts that do occur by identifying some good that will be realised, or some harm avoided. In section 4.7 I will therefore discuss how Persson and Savulescu take the first approach, and how they challenge the claims that moral bioenhancement will impact moral autonomy. A frequent argument provided by supporters of moral bioenhancement is that we do not regard the most virtuous among us as less free on account of their tendency to mostly do the morally right thing. The argument would then go on to claim that, in a similar manner, we shouldn't regard morally bioenhanced individuals as less free on account of their being less likely to perform morally harmful acts. This distinction will be discussed in detail in section 4.7, in order to illustrate that the matter is not as straightforward as the proponents imply.

In section 4.8 I will address Persson and Savulescu's arguments regarding the second approach, mentioned above, namely, the moral justifiability of impacts on moral autonomy, should they occur. Here, their claim that well-being and safety must be balanced with the requirements of autonomy will be briefly discussed. This will be followed by a discussion of various thought experiments that have been presented by the proponents of moral bioenhancement, the responses of opponents, and some insights from commentators on the issue of moral responsibility in general. These thought experiments are designed to illustrate the view that freedom does not necessarily require that one possesses alternative possibilities regarding action, a view that if true, would undermine Harris' view that morality requires freedom to fall. In section 4.8, I will also discuss Ronald Dworkin's interpretation of John Stuart Mill's distinction between liberty as licence and liberty as independence. Harris utilises this distinction to illustrate what is actually at stake in the moral bioenhancement debate. This distinction also merits discussion as Mill's ideas may be appropriated by both sides of the debate, depending upon how liberty is interpreted. I will then conclude the section with Persson and Savulescu's interpretation of autonomy, a distinction that they make between subjective and objective alternatives for action, and the relevance this has for their understanding of autonomy.

Section 4.9 includes a debate between DeGrazia, a proponent of moral bioenhancement, and Harris. DeGrazia argues that moral bioenhancement would meet three conditions required for autonomous action, to which Harris responds by arguing that whilst it would seem to be the case, the conditions would only be met due to the changes wrought by moral bioenhancement to fundamental dispositions. To conclude the chapter, in section 4.10, I will present some other interpretations of freedom that have been identified in the literature. Bublitz has identified the freedom of action and freedom of will as the two types of freedom that are engaged with by both proponents and opponents of moral bioenhancement. There is, however, a third type of freedom, namely, the freedom of mind. Bublitz argues that this type of freedom may well be impacted upon by moral bioenhancement and thus further investigation is required.

4.5 Overview of the autonomy debate

While Persson and Savulescu are the dominant proponents of moral bioenhancement, Douglas was the first thinker to address the possibility of biologically enhancing our morality[46]. In his seminal

[46] In a later paper, Douglas points out that his discussions of moral bioenhancement are not of pragmatic value, as he is not of the view that many individuals would willingly consent to be morally bioenhanced (2013:161). Rather, he engages with the problematic in order to emphasise his disagreement with arguments that view all forms of moral bioenhancement as ethically problematic, in principle.

article on the subject, Douglas anticipates potential concerns regarding the fact that the enhancement of motives – his interpretation of what moral bioenhancement should target – could remove the ability to have, and act from, morally bad motives (Douglas, 2008:239). However, depending on the way in which freedom is interpreted, Douglas argues that moral bioenhancement could, in fact, result in increases of freedom. Regarding the concern for freedom, he concedes that while we may not see much worth in acting from morally bad motives, "the *freedom* to hold and act upon them is valuable" (Douglas, 2008:239). However, he argues that if an individual has freely consented to moral bioenhancement, then the loss of this freedom would not, necessarily, be morally problematic. Furthermore, he points out that the concern that moral bioenhancement will lessen or remove the freedom to act in accordance with morally questionable motives is based upon a particular conception of freedom as requiring "the absence of [both] external…[and] internal psychological constraints" (Douglas, 2008:240). The concern is that while the actions of individuals who have freely consented to moral bioenhancement may be free from external constraints, they would, nevertheless, be subject to internal restrictions.

Douglas posits that such a conception is generally associated with the view that the self consists of a "true or authentic self, and a brute self that is external to this true self" (2008:240). The brute self would presumably consist of instinctual and emotional drives and urges. Thus, an increase in freedom would be associated with permitting the true self to be unhindered from the influences of the emotionally driven brute self. The concern, according to this conception, however, is that moral bioenhancement in affecting emotional responses would alter not only the brute self, but would, in some way, impact or subdue the true self. However, Douglas disagrees with this interpretation, arguing that his conception of moral bioenhancement which would be directed towards *lessening* the force of counter-moral emotions, such as racial aversion towards out-groups, would produce the opposite effect. It would work to enlarge the freedom of the true self due to the fact that it would lessen the force of the emotions associated with the brute self.

Thus, moral bioenhancement could be regarded as a means of enlarging the "freedom to have and to act upon good motives…[rather than interpreted as means of lessening] freedom to have and to act upon bad ones" (Douglas, 2008:240). However, while Douglas' response may be relevant for those who do make such distinctions between a true and brute self, it is not self-evident that those who have a concern for the way in which moral bioenhancement will produce internal psychological constraints base their concerns on such a distinction. Furthermore, Douglas' response aside, such a distinction would be a problematic basis upon which to launch an argument

against moral bioenhancement, in the absence of a convincing account of human psychology and selfhood. I would argue however, that there is merit in a more substantial account of the intuition that this distinction captures and will therefore develop this line of argumentation in detail in chapter 5.

In the same year as Douglas' first publication on the subject, Persson and Savulescu presented their arguments for moral bioenhancement, introducing the ideas that would come to dominate the debate. As mentioned in chapter 1, in their seminal article they argue that cognitive enhancement presents an enormous risk to humanity as it will equip individuals with greater capabilities to exact massive harms on a global scale, potentially leading to ultimate harm. Due to this risk, they conclude that cognitive enhancement should only be permitted if it occurs alongside compulsory moral bioenhancement (Persson & Savulescu, 2008:162). While cognitive abilities are, of course, essential for moral judgements, they posit that there is more to morality than rational deliberation. The most pertinent claim that Persson and Savulescu make in this first publication on the issue, is that, contra Socrates, and in agreement with Douglas, morality requires not only finding out or knowing what the good is, but more importantly, being sufficiently motivated to act upon it (2008:173). In the absence of the will, or motivation, to do what is right, they argue that morality is empty. They see this lack of motivation as the major moral problem of the 21st century. Addressing it, they posit, would 'solve' a variety of problems facing humanity.

As mentioned in previous chapters, Harris is renowned as a stanch supporter of biotechnological enhancement, and, in particular, of cognitive enhancement. He has, however, critiqued moral bioenhancement, as presented by its supporters, on the grounds that it would not amount to a form of moral enhancement due to the fact that it would negatively impact upon our moral autonomy, where the latter is interpreted in terms of our ability to perform truly moral actions, rather than acting in a way that simply has moral consequences. In light of this, Harris argues that the most reliable, comprehensive and safest forms of moral enhancement are the stalwart techniques that have long served humanity. These include traditional mechanisms such as "socialization, education and parental supervision" (Harris, 2011:102). In his first critique of Persson and Savulescu's position in 2011, Harris presents an interesting approach to the issue of moral autonomy and the impact that moral bioenhancement could have on it. He utilises a description from John Milton's majestic, extended narrative poem *Paradise Lost* in which God is described as having "made…[man] just and right, sufficient to have stood, though free to fall" (Milton, 1667, in Harris, 2011:103). In other words, Milton argues that human beings were created in such a way

that when faced with a choice between right and wrong, they possess sufficient moral capabilities to make the right choice; but, because they have freedom of choice, they also possess the ability to make the wrong choice.

One of the crucial points of dispute in the moral bioenhancement debate hinges on this supposed 'sufficiency of humankind to stand'. Harris does concede that the Miltonian claim that human beings were made "just and right" (in Harris, 2011:103) is a "vainglorious" (ibid.) one. However, he also subtly contradicts himself when he argues that if we think of ourselves as the products, not of divine creation, but rather, as the outcome of the process of evolution, then it is clear that "we have certainly evolved to have a vigorous sense of justice and right, that is, with a virtuous sense of morality" (Harris, 2011:103). However, Persson and Savulescu would dispute this conclusion. As discussed in chapter 1, the argument that they put forward, in multiple publications, is that our evolved moral capacities are not up to the task of addressing the problems that currently face humanity. Their arguments in this regard are particularly convincing and this issue is therefore clearly a major point of contention in the debate which warrants further discussion than Harris has provided. In other words, if Harris thinks that our morality is up to the task of addressing the problems that Persson and Savulescu outline, then he must provide more of an argument to support this position.

With his use of Milton's creation story, Harris is not attempting to provide a religious justification for his argument against moral bioenhancement. Rather, he refers to this formulation because in his view it encapsulates, in the pithiest form, just what is at stake in the moral bioenhancement debate (Harris, 2011:102). Milton's sentiment regarding this matter is, however, not novel. It is a long-held intuition, commonly known in theology as the free will defence. The free will defence is a theodicy that is offered as an answer to the problem of evil. The problem of evil has long served as a form of argumentation, or 'proof', against belief in the theistic God. It points to the immense evil and suffering perpetrated, and experienced, in the world, and argues that if God were truly omnipotent, he would be *able* to intervene and prevent, if not all, but some of the more severe atrocities; and, if he was truly good, he would have the *desire* to intervene and protect his creations from the worst forms of suffering. The fact that evil persists in the world implies that either God is not omnipotent, or that he is not truly good. This leads to the conclusion that it is likely that the theistic God does not exist.

A theodicy, such as the free will defence, is a form of argument that aims to protect religious belief, as its goal is to somehow explain or reconcile the presence of evil in the world with the existence of the theistic God. The Free will defence posits that in order to create true human agents, rather than puppets, God had to create us with free will, with the capability of choosing how we will act. However, in doing so, there was the risk that we subsequently possessed the freedom to do wrong, thereby bringing suffering and evil into the world. Evil, according to this view, thus originates due to the requirement that for human beings to be truly human, they must be free to do wrong. The deeper intuition that supports this argument is the view that true moral agents are of more intrinsic value than non-agents who are compelled to act in a particular manner. This intuition can also be reformulated as the view that the choice to act in a morally preferable manner has more value if it is made in the face of temptation and the possibility that one could equally choose to do the wrong thing. These are deep-rooted, long-standing human intuitions regarding moral autonomy that are difficult to challenge and they provide impetus to Harris' argument. However, they also serve to obfuscate some of the relatively unsupported claims that Harris makes, such as whether or not moral bioenhancement would, in fact, subvert moral autonomy as decisively as he implies.

Returning to Milton's description of the requirements of moral autonomy, Harris argues that it perfectly encapsulates "the human condition and…the precious nature of freedom and in particular free will" (2011:103). The nature of autonomy "requires not only the possibility of falling but the freedom to choose to fall" (ibid.). Harris argues that the kinds of purported moral bioenhancements currently discussed in the literature are "unlikely to leave us…free to fall" (2011:104). As mentioned in previous chapters, Harris is a rationalist, and this clearly informs his view on the nature of morality. Contrary to Persson and Savulescu, he argues that moral acumen does not involve "'being better at *being* good', rather it is being better at *knowing* the good and understanding what is likely to conduce to the good" (ibid.). Furthermore, Harris posits that freedom lies in "the space between knowing the good and doing the good" (ibid.). While Persson and Savulescu's focus is on how the biological enhancement of motivation could close this space, Harris implies that the traditional ways of attempting to close this space are preferable, due to the fact that because they involve rational deliberation, rather than compulsion, they are freedom-preserving.

While Harris associates *doing* good with *knowing* what is good, he doesn't hold an entirely Socratic view regarding morality, as he acknowledges the existence of the above-mentioned, sometimes unpredictable, gap between knowing and doing what is good. He briefly explores possible

explanations for why this gap exists, describing it as possibly indicative of the presence of a weakness of will, the term identified by the ancient Greeks as *akrasia* (Harris, 2011:104). Akrasia occurs when one is fully cognisant of what the morally correct – or preferable, in a non-moral sense, as akrasia afflicts not only moral action – course of action would be; but, for various, seemingly inexplicable reasons, one refrains from enacting it. As Akrasia is thus somewhat mysterious, as it cannot simply be attributed to the presence of competing options, temptations or simply a weak moral character; it may be rooted in complex aspects of human psychology[47]. However, the importance of attempting to more adequately understand this realm in which akrasia operates, which is one and the same space that Harris refers to above, is paramount for the moral bioenhancement problematic and thus cannot be glossed over.

As recognised by Rakić, and mentioned in section 2.3.1, this "discrepancy between what we do and what we believe is right to do might be the greatest predicament of our existence as moral beings" (2012:120). Clearly – and despite Harris' optimism regarding the matter – there are many thinkers who are of the view that this area of our moral development remains lacking. However, for Harris, regardless of whether or not we may fail miserably at closing this gap between knowing and doing the good, our moral freedom that lies in this gap is paramount as it is the prerequisite for morality itself. Harris' concern is that moral bioenhancement would be akin to moral compulsion. He argues that "without the freedom to fall, good cannot be a choice; and freedom disappears and along with it virtue. There is no virtue in doing what you must" (Harris, 2011:204).

4.6 The emotion/reason dichotomy revisited

Harris reaffirms his rationalist proclivities in his discussion of Douglas' argument for the mitigation of counter-moral emotions as a form of moral bioenhancement. In particular, he examines the claim that a potential candidate for moral bioenhancement, in this regard, would be the impulse towards racial aversion. Harris is of the view that beliefs, such as those that inform racial aversion, are not emotionally rooted but are rather cognitive in nature. In other words, he regards such beliefs as based upon incorrect factual content regarding the groups in question. Thus, directly modulating emotions would simply sidestep these false beliefs and biases, or the faulty process of deliberation that may have informed such beliefs, rather than correcting them. In this regard, he argues that the most "obvious countermeasure to [such] false beliefs and prejudices is a combination of rationality

[47] There are a variety of conflicting theories that attempt to explain the sources of akrasia. See Mele, 1983; Davidson, 1980; Ainslie, 2001.

and education" (Harris, 2011:105). Here, he includes cognitive enhancement as a possibly helpful tool.

For Harris, a moral action that occurs without any, or sufficient, rational deliberation – as he posits would be the case with an intervention that seemingly bypasses this process – is not a moral action at all. Of course, a possible response to Harris' concern that rational deliberation would be bypassed would be to point out that in the case of voluntary moral bioenhancement, the decision to undergo moral bioenhancement would presumably be due to having recognised that one is, in some way, in need of moral bioenhancement to begin with. Thus, agreeing to moral bioenhancement in the first place would indicate that, in all likelihood, this decision would have been informed by a process of deliberation, namely, the realisation of the possible falsity or problematic nature of certain beliefs that one has, such as those informing any racial biases, despite the grip that they exert on the individual, and the desire to change these beliefs.

This response aside, as mentioned in previous chapters, for Harris, mitigating counter-moral emotions does not constitute genuine moral enhancement. His arguments elsewhere are interesting in pinpointing just how he thinks moral bioenhancement would erode human morality, by compromising our moral autonomy. Chan and Harris discuss the work of the neuroscientist Molly Crockett who, as mentioned in section 2.4.2 of chapter 2, has investigated the effects of serotonin on moral judgement (2011). Crockett's experiments show that administering serotonin to subjects increases both "pro-social behaviour" (Chan & Harris, 2011:130) and their disinclination to effect harm upon others. Ostensibly, an intervention that increases cooperation between individuals and results in them being less likely to wish to harm each other may appear to be an obvious moral enhancement. However, while Chan and Harris view such an intervention as influencing behaviour, they do not see the resultant behaviour as "moral behaviour" (2011:130). They argue that such behaviour may have moral consequences, however, this is not sufficient for it being moral itself.

The primary problem that Chan and Harris have with viewing SSRIs as potential moral enhancers is that they argue that what SSRIs actually do is heighten the relevant *feelings* or *emotions*, such as empathy, associated with pro-sociability (2011:130). The reasons provided for any behavioural changes brought about due to such interventions would then have little genuine cognitive content. Rather, they would track the intensity of feelings, or emotional response, to a possible course of action. Although Chan and Harris do admit that research in the field of neuroscience indicates that

moral responses are a product of both affective and cognitive components, they nevertheless conclude that "of the two, it must be reasoning that pulls in the direction of morality" (2011:130). When an individual makes a moral judgement, she must generally adjudicate between competing options. She may, of course, assess these options through examining her emotional responses to the various possible courses of action. However, to actually make her final decision, she must be able to deliberate the reliability of her emotional responses in terms of how various potential courses of action support other important morally relevant factors. Examples of such factors include whether or not the potential courses of action support "principles…values…and…moral objectives" (Chan & Harris, 2011:130) that she identifies with, her moral code in other words. Furthermore, she will then assess all of the above in terms of relevant contextual factors through the "use [of] moral reasoning" (ibid.)

Chan and Harris then conclude by arguing that in cases where serotonin has been administered, with the effect of, for example, increasing pro-sociability or harm aversion, these increases will somehow thwart, or weaken, the cognitive components of the process. In other words, they will interfere with the individual's application of her pre-existing moral code to competing options as well as her ability to weigh up and assess the potential courses of action against relevant contextual factors. Chan and Harris are of the view that in such cases, the individual will be *compelled*, presumably due to the strength of her emotional responses, to ignore the above-mentioned cognitive and deliberative component of the process. Rather, they posit that under the influence of serotonin, the individual's decision will be the product of a second check of her emotional responses, with the final choice based upon that decision which elicits the strongest emotional reaction.

As mentioned in chapter 3, they argue that this is akin to Wittgenstein's example of buying a second copy of the same newspaper to verify what one has read in the first newspaper. While the individual may believe that she is making a free moral choice, the intensity of emotional reaction will render a truly freely made moral choice impossible. Thus, the overriding of reason by emotion is something that Harris seems to think will be an inevitable result of moral bioenhancement. In a later publication, he asks "if the good involves feeling the right way, how do we know that we are feeling the right way?" (Harris, 2013b:171). To express this idea in a slightly different manner, Harris implies that the moral decisions that inform our actions must rely upon something other than our emotional states which are predominantly internal states. There must be interplay between these internal states and an external reality, with relevant contextual factors taken into

consideration, otherwise decisions will be based upon a reinforcing internal loop of emotion, with no way of assessing the legitimacy of the latter.

Harris is correct in his argument that morality should never be solely a matter of emotional response[48] due to the fact that our emotions are frequently informed by prejudices, unquestioned assumptions and a variety of other arbitrary influences. However, one must inquire as to whether serotonin, or any other SSRI, for that matter, would, in fact, produce the effects that Chan and Harris fear, if used for enhancement rather than therapeutic purposes. It seems doubtful in light of the millions of individuals who currently take SSRIs to treat conditions such as depression and anxiety, without the seemingly obvious bypassing of their rational deliberative capacities. Furthermore, as discussed in section 3.2 of chapter 3, Harris has been criticised for not paying adequate attention to neuroscientific findings regarding the way in which moral decisions are actually made and the important role of affective components in this regard (Crockett et al., 2010:E184). As pointed out by Persson and Savulescu, findings in the field of neuroscience are beginning to shed light on long-held ethics disputes. One such dispute that has been debunked, was mentioned in section 2.4.2 of chapter 2, namely, the position that views Kantian ethics as primarily reason-based, while consequentialist judgements are viewed as predominantly driven by affective components. Research has illustrated that the opposite is the case. Experiments utilising functional magnetic resonance imagining (fMRI) of the brain, showed that brain areas associated with emotion were far more active when making deontological, supposedly reason-based, judgements than when making consequentialist or utilitarian ones, which elicit a more calculative or rational response (Greene et al., 2001:2106).

Furthermore, as Singer discusses, generally when we provide what we think to be legitimate reasons for a morally relevant opinion that we hold, they are likely to be post hoc rationalisations for what is actually an "initial intuitive [or gut] response" (2005:350), that is, in all likelihood, informed by evolutionary origins. Haidt has described this as a situation in which "the emotional dog…[is wagging its] rational tail" (2001:814). One way in which this is illustrated is discussed by Singer who argues that when faced with a course of action we find morally abhorrent, such as an account of a once-off occurrence of incest involving mutually consenting adults, we are very skilled at providing practical and seemingly intelligible reasons to illustrate why it is wrong

[48] But to be fair to Persson and Savulescu they have never argued for this position. They have frequently mentioned that rational deliberation is a necessary component of moral decision-making. However, presumably by identifying emotions as the target of moral bioenhancement, Harris views them as simply paying lip service to the importance of rational deliberation in moral decision-making.

(2005:337). We may, for example, argue that it is wrong because a child might be conceived or one of the individuals may feel uncomfortable with the act. However, even if each of these practical concerns are decisively debunked, we are told that contraception is being utilised and that there is genuine consent by both parties involved, we will generally persist with our opinion that a practice such as this is just simply wrong. Here is a clear example of a strongly held belief that is based upon an intuition or taboo which has evolutionary origins. The relevance of the above points for Harris' argument, is, that he is, in all likelihood, underestimating the role that emotional responses already play in our moral decision-making and, in turn, overestimating the role played by rationality, or at least a detached Kantian type of rationality, in this process[49]. Of course, if this is true, Harris could then respond by arguing that it supports his argument that what is actually needed is cognitive, rather than emotional or moral, bioenhancement.

4.7 Will moral bioenhancement impair moral autonomy?

Persson and Savulescu have responded in a number of publications to Harris' initial critique of their argument, specifically his use of Milton's conception of freedom. They posit that Harris has exaggerated the effects that moral bioenhancement would have on moral decision-making, describing his concerns as "extreme, perhaps hyperbolic" (Savulescu & Persson, 2012:403). They argue that this is specifically true of his fear that moral bioenhancement would somehow eradicate our moral autonomy, by rendering us simply unable to do wrong. As alluded to above, there are two ways of responding to concerns such as those that Harris puts forward. Firstly, one can investigate the nature of moral decision-making in order to work out whether the type of moral bioenhancement interventions that are proposed would, in fact, have the effects that Harris fears. Secondly, one can take an entirely different, albeit more controversial, approach and ask a different kind of question. If moral bioenhancement did, in fact, produce the effects that Harris fears, would this "be a bad thing all things considered?" (Savulescu & Persson, 2012:402). Savulescu and Persson investigate both options.

In terms of the first way of responding to the concern for moral autonomy, Persson and Savulescu discuss the nature of moral action, pointing out that one's view of what is considered a morally

[49] This should not imply moral scepticism, as Singer points out. In other words, it is not that we must accept the position that what we take to be our moral deliberations are simply the product of "emotionally based intuitive responses" (Singer, 2005:351) justified by reasons and arguments constructed after the fact. Rather, it is to acknowledge the role play by emotions in our moral deliberations. Furthermore, we could attempt to distinguish those "moral judgements that we owe to our evolutionary and cultural history, from those that have a rational basis" (ibid.). Singer argues that the "axioms" of utilitarianism would be an example of the latter as the admonishment to hold the happiness of all human beings as equally important would have been at odds with the evolutionary mechanism of natural selection to observe altruism towards only one's kin.

good act is directly informed by the moral theory one espouses (Savulescu & Persson, 2012:403). For example, when faced with a decision that has moral relevance, a utilitarian would choose the action that produces the most favourable long-term consequences for all individuals affected. This is decided by balancing the utility produced by an action over any potential suffering that may result. Furthermore, utilitarianism operates from a foundation of absolute equality whereby each individual's utility must be given commensurate weight. With this interpretation of what is considered morally good, a utilitarian would approach a project of moral bioenhancement in a particular manner. Such an enhancement would firstly include cognitive enhancement in order to be able to make a more reliable analysis of the potential consequences produced by an action. Secondly, it would require improvements in "impulse control" (ibid.) in order to ensure that one is sufficiently motivated to perform the act that is considered to be morally correct. Thirdly, and most important for utilitarianism, an intervention would have to assist in boosting the disposition of selflessness, as the action considered to be the most morally optimal could be one that is less beneficial, and thus less desirable, for the individual concerned.

As discussed in previous chapters, Savulescu and Persson associate this quality of selflessness, or ability to consider the interests of others, with a sense of altruism, one of the dispositions they target for moral bioenhancement. Furthermore, as Savulescu and Persson point out, the requirement to set aside one's own interests in order to consider the wellbeing of others is not specific to utilitarianism, but is rather a dictate of morality in general (2012:403). Of course, moral theories do differ regarding the degree of observance they place on this requirement for selflessness or altruism; with utilitarianism being considered the most demanding moral theory in this regard (Savulescu and Persson, 2012:403)[50]. Nevertheless, Savulescu and Persson argue that "increasing the willingness to sacrifice one's own interests for the benefit of others is a moral enhancement, on any account of morality" (2012:404). Focusing on a particular trait or moral disposition that is universal to all accounts of moral action, as a target for moral bioenhancement, in the way that Savulescu and Persson do, helps them avoid the thorny meta-ethical issue of having to justify a particular moral theory.

They then look at the question of whether increasing our levels of altruism would somehow produce the effects that concern Harris. In other words, if we were morally bioenhanced to, for example, be more selfless and compassionate towards others and to make sacrifices that may, to a certain

[50] This point was discussed in the conclusion of chapter 2 when I provided a working definition for moral bioenhancement that includes a reference to empathy, interpreted as closely associated with the quality of selflessness.

extent, impinge upon our own degree of comfort but would produce favourable aggregate level improvements, would this compromise something intrinsically valuable to us, namely, our morality? To rephrase the question, would making ourselves more altruistic, than we otherwise would be if no intervention had taken place, somehow compromise our moral autonomy? Would we still have the choice to decide whether or not to act in a particular manner; in other words, to choose not to be altruistic?

It seems that for Harris to be satisfied that our moral autonomy remains intact, the choice to not be altruistic in any given situation would have to be preserved. However, if this choice is truly preserved, and individuals could easily choose to do wrong after having been morally bioenhanced, or at least choose to not act upon their boosted altruism, we would have to ask if moral bioenhancement had actually taken place. It seems that, at the very least, to be truly considered as a moral bioenhancement, an intervention would have to increase the *probability* that individuals will act upon the enhanced dispositions. This problem lies at the heart of the matter. As has already been mentioned in previous chapters, Persson and Savulescu are quick to use particular examples to counter these concerns. They point out that we do not consider women less morally free than men because they are generally predisposed to exhibit more empathy than men (Persson & Savulescu, 2012:112)[51]. Furthermore, we do not consider individuals who display high levels of altruism and compassion for others to be less morally free than the majority of individuals who place a premium upon their own levels of wellbeing over the wellbeing of others (Persson & Savulescu, 2012:112)

Persson and Savulescu argue that in the case of the enhancement of empathy, it is unlikely that individuals would be overwhelmed with emotion, thereby being reduced to the status of automata (Persson & Savulescu, 2013:128). Rather, it could be the case that after such an enhancement, they would simply be acting "for the same reasons as those of us who are most moral today do, and the sense in which it is 'impossible' that they do what they regard as immoral will be the same for the morally enhanced as for the garden-variety virtuous person: it is psychologically or motivationally 'impossible'" (Persson & Savulescu, 2013:128). Thus, they are of the view that any loss of freedom produced by moral bioenhancement will be similar in kind to the supposed lack of freedom that truly virtuous individuals experience in their feelings that it would be impossible to commit certain

[51] Whether or not the tendency to be more empathetic is a product of a biological predisposition or socialisation is beside the point. Rather, what is important here is that given that we agree that women are in general more empathetic than men, do we regard them as less morally free in some way.

immoral acts. However, they argue that we would never regard such individuals as lacking in freedom (Savulescu, Douglas & Persson, 2014:101).

Harris has argued, however, that *freedom to fall* is not akin to possessing the "ability to act otherwise" (2014:373), as some of his critics imply. He agrees with Persson and Savulescu (2014) and DeGrazia (2014) that I may consider myself unable to murder someone I love due to the fact that I have inculcated the belief that killing someone who is innocent and whom I love is wrong. In such a situation, I am incapable of acting otherwise, as killing an innocent or loved one is an impossibility for me; it is simply not an option for action. Freedom to fall, on the other hand entails my possessing "the freedom to decide whether or not to fall for reasons, which have to do with what is best 'all things considered'" (Harris, 2014:373). For Harris, moral bioenhancement would threaten this ability to make *all things considered judgements*, and thus, would threaten moral autonomy. What he means by this is that rather than acting upon reasons or a process of moral deliberation, we will act according to the intensity of our emotional responses, which are not always reliable, and more importantly, for Harris, are simply not the substance of morality.

However, various examples have been discussed by the proponents of moral bioenhancement to illustrate why Harris is mistaken in this regard. Savulescu and Persson discuss an individual who is concerned for how little he is moved by the plight of the poor, and in particular, a beggar he encounters on a daily basis (2012:407). Upon taking a moral enhancement drug that boosts his altruism levels, he subsequently *feels* more compassion for others, which leads him to have a keener sense of how it would be to have nothing and to have to rely on others. However, the individual has a pre-existing general belief that the best option is not to give money to people who beg as they may use it to purchase alcohol or drugs which would be detrimental to their health. Therefore, based upon this belief, which is a product of his having morally deliberated, the individual rather decides to buy food for the beggar.

Savulescu and Persson see this example as indicating that Harris' fears are unfounded. They posit that the moral enhancement drug that was taken, acted "like a pair of 'moral spectacles' clarifying his vision of the other" (2012:407). Rather than acting mindlessly in the way that Harris fears, the individual has still employed his deliberative capacities and acted upon a previously held cognitive belief regarding the most effective way of helping those in need. In other words, he has legitimate reasons for his action; what has changed, is that because he is now able to empathise more, this makes it harder for him to simply ignore the beggar and do nothing. Another way of articulating

this point is that by boosting the feeling of sympathy, we are increasing "the probability that we do what we believe that we ought" (Persson & Savulescu, 2016:264). Persson and Savulescu see an increase in the likelihood of individuals doing what they morally should do as an obvious example of a moral enhancement (2016:264)[52].

While it is true that we wouldn't necessarily regard the individual in Persson and Savulescu's example as lacking in freedom, their claim is only true for this kind of example. In other words, it is true only for cases in which individuals have pre-existing beliefs that may not be strong enough to motivate action, but are nevertheless present. In their example, Persson and Savulescu describe the individual as being concerned by the lack of empathy he feels for the poor. Thus, he is acting upon a belief that he already has, albeit one that possesses insufficient motivational force. He has problematised his lack of empathy enough to undergo moral bioenhancement, thus implying that he regards experiencing empathy for the poor as a good and worthwhile disposition to have. There is, however, a different scenario that Persson and Savulescu have not accounted for. If, before being morally bioenhanced, an individual did *not* regard helping the poor as a moral obligation, or simply as a good and worthwhile thing to do, but then, after being morally bioenhanced, for unrelated reasons, and on the basis of increases in empathy, felt compelled to assist them, the process of reasoning would be entirely different in comparison to an individual who assists the poor due to a weak, but pre-existing belief. Thus, Persson and Savulescu's argument that Harris' fears regarding compulsion are unfounded, holds for those who have a weak belief that helping the poor is a worthwhile endeavour, where such a belief is not strong enough to motivate them. However, in the case of an enhanced individual who has never possessed such a belief, and who then comes to the realisation that it is right to help others who are less fortunate, purely as a result of an intervention, this would be a different matter. This is an important point that I will develop further in chapter 5b.

Returning to the point that Harris makes regarding the ability to make all things considered judgements, Savulescu and Persson's discussion of their example glosses over another important point that may be made regarding sympathy. Harris argues that a feeling of sympathy for the poor

[52] Of course, whether this would be true would depend on contextual factors. History is littered with examples of atrocities committed by individuals who believe that they are doing what they morally ought to do. Therefore, this interpretation of moral enhancement requires further substantive content. For example, to be considered as a moral enhancement, an intervention would have to increase the likelihood that individuals will do what they morally ought to do, in accordance with those moral principles most likely to have universal support. As mentioned in chapter 2, they do qualify their definition in a later publication, by adding that acting "morally…[occurs] when one does the right thing, and for the right reason(s)." (Savulescu, Douglas & Persson, 2014:95). They also admit that this qualification would be dependent on what one's interpretation is regarding "right action and right motivation" (ibid.)

should not be the source of my deliberating that I should thus assist them in some way (Harris, 2013b:172). Rather, it is my considered judgement that the poor require assistance, and, that regardless of my feelings on the subject, it is the right thing to provide assistance to them, that should elicit feelings of sympathy (2013b:172). In other words, sympathy comes after the fact.

Furthermore, Savulescu and Persson's example fails to adequately engage with some other important points that should be addressed. Firstly, it seems that reason-responsiveness is only part of what makes an action morally autonomous. In the case of an individual who was opposed to giving help to the poor before his enhancement, he would presumably be able to come up with reasons to explain his changed behaviour after the enhancement. However, whereas before the enhancement, his reasons would have explained his opposition to giving help to the poor, he would now have different reasons. Furthermore, this change in his reasons would be *entirely* due to the enhancement and would therefore be a case of post hoc rationalisation. Before the enhancement, the individual may have justified his decision to ignore the beggar in a variety of ways. He could, for example, argue that giving help is pointless as it would make no discernible difference in the life of the beggar who will always require more help. Or, he could argue that his lack of empathy for the beggar is rational, as she could, in all likelihood, alleviate her situation and will be less likely to do so if she receives regular help from strangers. The important point here is that the enhanced individual in this example would never have possessed the belief, or come to the realisation, that it is right, to help others who are less fortunate and this is perhaps what perturbs Harris. It is not simply the presence of some, or any, cognitive content that is required for an action to be moral in the sense that Harris interprets it, but rather, an action should be motivated by the 'correct', or appropriate, moral belief, where correct and appropriate imply the most reasonable belief.

While it could be argued that whichever belief the individual possesses, and uses to justify his lack of giving help to the beggar, is already the product of an emotional response, this is not the entire story. On the one hand, beliefs are not entirely socially constructed; they are also influenced by temperament or psychological dispositions in general. In our story, the individual could have low levels of empathy. On the other hand, and as highlighted in the discussion of narrative identity, given knowledge of the individual's life story – and absent of any psychopathology – we would be able to provide a narrative that would explain his beliefs and link them to discernible sources, even if this could not be done perfectly. Possible contenders for the sources of such beliefs would be vast and complex. An obvious example would be the nature of the moral education that he had

received, including the behaviour of role models in his upbringing, and the extent to which he internalised this behaviour. The important point, as stated above, is that after the moral enhancement, he would now possess new reasons for his change in behaviour. These may be entirely legitimate reasons, such as a new belief that it is good to help those who are less fortunate, or, that one would want to receive similar help if one was in such a position. However, the actual reason for any change in his beliefs would be the administering of the moral enhancement.

It must be noted that the fact that the individual has experienced a change in the reasons he uses to justify his behaviour is not problematic in itself. Whilst changes in deeper and more enduring moral beliefs[53] are rarer, individuals are inured to frequent changes in their interpretations of phenomena. The latter generally occurs when more information is acquired, thus leading to the individual possessing a more comprehensive overview of a situation or phenomenon. However, along with any new beliefs and concomitant reasons for these beliefs, regarding one's new stance towards assisting those who are less fortunate, one would have to admit an additional reason regarding why one's beliefs have changed. This additional reason would be associated with the knowledge that one has undergone a moral bioenhancement.

Thus, Harris seems to be getting at something with his concern that moral bioenhancement would impact or alter morality, as we conceive it, in some way. It is not that that moral bioenhancements would result in compulsive behaviour that is devoid of rational explanations. Rather, the concern is for the way in which moral bioenhancement would influence how we come to change our beliefs regarding right or wrong, that motivate subsequent behaviour, that seems to be an important matter. There does seem to be a non-trivial difference between a change in beliefs due to having acquired additional contextual information, or, having experienced a paradigm shifting event, and a change in belief due to a biological intervention. However, I would argue that where there is a weak pre-existing belief that is strengthened by moral bioenhancement, this would not necessarily be problematic, whereas in cases in which the process of moral bioenhancement supplanted one belief with an opposing one, this would be cause for concern. I will elucidate this matter further in chapter 5b.

[53] By deeper and more enduring beliefs, I mean beliefs that are considered important to an individual, regarding what she considers to be right and wrong, and that are relatively impervious to major change without a vast upheaval in personal identity.

4.8 Are impacts on moral autonomy morally justifiable?

As mentioned at the beginning of section 4.7, there is a second possible way of responding to the claim that moral bioenhancement will threaten moral autonomy. One could argue that even if moral bioenhancement did, in fact, produce the effects that Harris fears, this could be morally justifiable. Here, Savulescu and Persson introduce their well-known, and highly controversial, hypothesis, the God machine. This hypothesis merits discussion, not on practical grounds, but due to the theoretical insights it affords into the issue of moral bioenhancement and moral autonomy. In other words, while it is a highly implausible idea that would be more appropriate in a dystopic novel, it is useful as it makes explicit the ideas underlying Savulescu and Persson's position.

4.8.1 The God machine

Savulescu and Persson imagine a futuristic "bioquantum computer" (2012:408) that is connected to the consciousness of all human beings and is able to somehow track "the thoughts, beliefs, desires and intentions of every human being" (2012:408). In the event that the machine perceives an intention to cause a serious harm to another individual, it is able to intervene instantly and alter this intention, unbeknown to the individual in question. In this way, the machine has eradicated serious interpersonal harms such as murder, rape and grievous bodily harm. While the machine does not intervene for smaller crimes or misdemeanours, there is subsequently less need for incarceration as greater harms simply cannot be perpetrated. Savulescu and Persson argue that in this futuristic hypothesis, morality is still possible, because if individuals opt to do good, or refrain from committing harm, they are doing so entirely freely as the machine only intervenes in the event that a decision has been made to commit a serious harm. They point out at that before the advent of the machine, punitive laws deterred people from performing harmful acts, and, in this regard, their freedom was not absolute. In fact, they posit that freedom has been increased due to the fact that fewer people languish in prison for extended periods of times. The difference between deterrence via punishment and the world of the God machine, is that it is now "literally impossible to do…[certain] things" (Savulescu & Persson, 2012:408). A person experiences the intervention as simply "chang[ing their] mind[s]" (Savulescu & Persson, 2012:408). Savulescu and Persson concede that the God machine wouldn't be a true moral enhancement as people could still have an intention to do wrong; their intention would simply be changed as soon as it was formed. However, they argue that so long as individuals had voluntarily signed up to the machine, autonomy would be preserved. In other words, voluntarily connecting to the machine would be a kind of "precommitment contract" (2012:409)

To explain the idea of a precommitment contract and illustrate how such a contract would be freedom preserving, Savulescu and Persson utilise the tale of Ulysses and the Sirens from Homer's *Odyssey* (Homer, 2004, in 2012:409-410). Ulysses was preparing to sail past the legendary region of the Sirens, whose voices were so bewitchingly beautiful that they hypnotised all men to go willingly to their deaths. He desired to hear this exquisite sound but wished to do so safely, and therefore, instructed his men to block their ears with wax to ensure that they wouldn't be able to hear the compelling, but deathly, siren songs. Ulysses, his ears unblocked, was bound to the mast, having given his men strict instructions to ignore his commands to release him. Once he heard the sirens, he predictably begged for his freedom. However, his men held steadfast to his first command to ignore him should he beg to be untied and he passed safely through the area.

Savulescu and Persson view this anecdote as effectively illustrating a situation in which an individual – after assessing the relevant factual information – makes a fully autonomous decision that will be binding, regarding a future state of affairs, in which he may later decide otherwise. Another example of such a scenario would be in individual who has recently ceased smoking and asks a friend to ensure that she refrains from smoking at a party. Here, the individual has formulated a considered desire to continue her abstinence from smoking, however, she foresees that her resolve may weaken in a future situation and puts a mechanism in place to ensure that her freedom will be curtailed should this happen. In other words, should she desire to smoke at the party, she recognises this will be a temporary desire that is at odds with her overarching desire to stop smoking. Both of these examples are indicative of what Savulescu and Persson describe as having an "obstructive or irrational desire which goes against…[a person's] best judgement" (2012:409). If we use these insights in terms of Savulescu and Persson's God machine example, the implication is that agreeing to connect to the God machine is congruent with having an overriding desire to not commit an act of interpersonal harm, and thus, being willing to surrender a portion of one's freedom to harm, should that desire arise later.

Harris argues, however, that the analogy is an inappropriate one to explicate their God Machine hypothesis. He argues that agreeing to hook yourself up to the God Machine would be akin to being free "to sell yourself into slavery" (Harris, 2016:106). Freely selling yourself into slavery is generally considered to be paradoxical. For Harris, freely hooking yourself up to the God Machine would be a similar paradox. Sparrow has also discussed this concern. He argues that freedom can be interpreted, as Philip Pettit has done, as requiring "non-domination" (1997:21 in Sparrow, 2014a:27). To illustrate this claim, he describes a slave who is under the dominion of a master,

who happens to be good-hearted. At any moment, the master may wield his power and thus interfere and control any, and every, aspect of the slave's life. The fact that the master refrains from doing so is merely a happy coincidence, as he could just as easily not have been good-hearted. As Pettit argues, most would agree that such a slave is not free; he only appears to be free. While, in the case of the example, he is able to lead his life as he sees fit, he nevertheless remains in the power of his master. He is, at all times, "subject to his [master's] power – regardless of whether or not he exercises it" (in Sparrow, 2014a:27). In a similar manner, and regardless of the fact that the God machine would only intervene in cases where serious harms are about to be committed, any individual hooked up to the machine would be under its power.

4.8.2 First and second-order desires

Harris and Sparrow's concerns aside, Savulescu and Persson's discussion of the possibility of having conflicting desires merits further discussion. In fact, this notion of conflicting desires was first discussed by Harry Frankfurt in his account of autonomy that has since come to be known as a hierarchical account of autonomy (1971). In attempting to define what confers autonomy, and thus, personhood, Frankfurt distinguishes between "first…[and] second-order desires" (1971:5-7). All sentient creatures have some form of will that drives them and can be understood as being constituted by their "desires and motives" (Frankfurt, 1971:6) which inform, or drive, the decisions that they make. These desires can be described as first-order desires, that, in their simplest description, are indicative of wanting "to do or not to do one thing or another" (Frankfurt, 1971:7). Humans, however, are unique in that they are seemingly the only creatures who, in addition to these first-order desires, are able to form opinions about their desires. This is by way of the uniquely human ability for self-reflection. Frankfurt calls these second-order desires or volitions (1971:7).

In other words, human beings are able to have desires, but, through self-awareness, they are also able to form preferences about their desires. They can realise that some of their desires are desires that they would rather not have. As Frankfurt succinctly puts it, human beings are "capable of wanting to be different, in their preferences and purposes from what they are" (1971:7). In terms of the above examples, the identifiable second order desires would be the desire to act morally and not harm others, to not succumb to the Sirens' call and thus to certain death, and to refrain from smoking. The more primal, first order desires or drives would be the momentary desire to commit harm or do wrong, to hear and act upon the Siren's song and to smoke at the party.

If we are to agree with Frankfurt and take second-order desires as constitutive of personhood, then, by implication, altering these desires would be problematic and inimical to autonomy, as self-determination, whereas altering first-order desires could be viewed as an enlargement of autonomy. In the case of moral bioenhancement, as pointed out by Hubbeling, we would have to ensure that it would not, in any way, alter second-order desires (2009:188). While Frankfurt's distinction makes intuitive sense, there are problems that plague such accounts. I will discuss this matter fully in chapter 5a, as despite the above-mentioned problems his account has major relevance for the issue at hand.

Returning to the God machine example, in the case of agreeing to hook oneself up to the machine, distinguishing between first and second-order desires would not be as straightforward a matter. This is because the second-order desire to hook yourself up to the machine, because you have a belief that it is wrong to cause harm, and thus, a desire to not commit harm, is not a straightforward matter. As Harris has pointed out elsewhere, harm is not a definitive notion; it is determined by contextual factors (2016:66). Thus, a seemingly fleeting first-order desire to commit a harm after one has been hooked up to the God machine would be deemed ethically permissible to be altered by the machine. However, this supposed first-order desire may in fact be constitutive of a second-order desire. In other words, whilst I may have a second-order desire to never commit a violent harm to another, there are contexts in which it could be necessary to act in such a manner in order to avert a worse harm from occurring[54]. In such situations, if the machine intervened and it became impossible to act and prevent the worse harm by performing a lesser harm, then my second-order desire would have been thwarted. Furthermore, this would, in all likelihood, have occurred without my knowing that it had happened.

4.8.3 Well-being and safety versus autonomy

Savulescu and Persson do admit that there would be some kind of loss of freedom in cases where the God machine did intervene, as, in these cases the individual's "moral identity" (2012:410) would have been subsumed into that of the God machine. However, they opt to then take the second approach, mentioned above, and argue that this loss of freedom would be offset by the great advantages of eradicating the world of serious harms. In other words, here they are contesting the

[54] An example of such a situation was discussed in footnote 33. I may be opposed to torture in-principle, however, when faced with the possibility of a bomb that will detonate imminently, killing thousands, I may condone torture to extract information from the individual who has planted the bomb and will not reveal its whereabouts. In the case of the God machine such a possibility could, of course, only arise if the individual who had planted the bomb was not hooked up to the God machine and I, the would-be torturer, was.

view that freedom is a fundamental right that should never be impinged upon. In support of this claim, Savulescu and Persson cite John Stuart Mill's renowned posit that "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not sufficient warrant" (Mill, 1863:23). In other words, according to this view, sometimes called *the harm principle*, self-government is highly valuable, and individuals should be left to lead autonomous lives, but only insofar as their autonomy does not impact on the well-being of others. As soon as the individual exercising of autonomy harms another, then it becomes illegitimate. Well-being and safety are therefore more important than autonomy, thus indicating that the latter is not an absolute value. Douglas shares this view, as evidenced by his claim that even if the modulations of emotions did produce the effects that thinkers such as Harris fear – thus lessening our ability to do wrong – this would not necessarily be a decisive argument against moral bioenhancement. He argues that "the freedom to do evil is less valuable than the evil is disvaluable" (Douglas 2013:166). His point is most certainly true, but it does not follow from its truth that moral bioenhancement is morally permissible!

Savulescu and Persson do not see their above-mentioned view as controversial, and, in the way that it was formulated by Mill, most would agree. Our freedom of action is not absolute; it is constrained by the very same freedom that others possess. However, in their use of the harm principle as a justification for the kind of freedom that would be lost in the case of the God machine, Savulescu and Persson do seem to be taking creative licence a step too far. The matter of the interpretation of the harm principle is also more complex than it *prima facie* seems. What is contested regarding the harm principle, is its degree of application. Depending upon the strength of one's utilitarian preclusions, this position can either be taken as a minimal or stand-alone requirement for moral life in its entirety, or as a foundational precept upon which more complex moral principles and human rights may be built. Savulescu and Persson are presumably of the opinion that Mill would support the former interpretation as he was of the view that the harm principle should govern, not because we have some right not to be harmed, but rather, because he would argue that a society in which the harm principle governs will be a preferable one – in terms of the maximization of utility – in comparison to a society in which it is absent. I will return to this matter further in section 4.8.5.

However, in terms of the relevance for the God machine, the matter is more complex than Savulescu and Persson's dealings with it implies. It may be true that an intervention which prevents

individuals from inflicting serious harms on others would be an unequivocally good thing. However, the crucial issue that they gloss over here, but are forced to address in a later publication in response to Harris, is the effect that the God machine would have on human morality itself. In other words, the ensuing impact on human morality could itself be regarded as a more serious harm to humanity than any of the individual instances of harms prevented by the machine[55]. The God machine is, of course, a scientific impossibility and Savulescu and Persson do not put it forward as a serious contender for moral bioenhancement. Furthermore, its main purpose, as stated by Savulescu, is to illustrate their point that "freedom has a limited – not unlimited – value" (Harris & Savulescu, 2015:19). However, it raises an important issue that must be addressed in order to explicate the concern for moral autonomy posed by moral bioenhancement in general.

4.8.4 Some other thought experiments

In a sense, the insights revealed by the God machine hypothesis are analogous to Robert Nozick's *Experience Machine* hypothesis[56]. The experience machine has frequently been utilised to show that utilitarianism fails to take cognisance of what truly matters to individuals. If the feeling of happiness, or the actual maximization of happiness, is all that matters to us, then there would be no good reason to fail to connect ourselves to the experience machine. However, Nozick is of the view that presumably most would not opt to connect themselves and this is because merely feeling a particular way is not all that matters to us. Rather, we wish to actually "*do* certain things, and not just have the experience of doing them…[and] we want to *be* a certain way, to be a certain sort of person" (Nozick, 1974:43). Nozick also draws the analogy between the experience machine and the imbibing of psychoactive drugs which produce effects of bliss and euphoria. In both examples, the experiences are artificial and there is a "lack of contact with deeper reality" (Nozick, 1974:44). For this very reason, drugs are frequently viewed as a means of escaping reality. These intuitions pinpoint something that is deeply important to most individuals, and thus, Nozick does not believe that many would connect themselves to the experience machine. For similar reasons, it is most

[55] Rakić, makes a similar point. He argues that if subverting our freedom – through compulsory moral bioenhancement – were the only way in which the human race could avoid extinction, then it would be preferable that the human race not survive. He argues that if we were to lose our moral freedom, we would risk losing "an essential element of…human existence, thus in a way getting already into the business of our self-annihilation" (Rakić, 2014:249). As mentioned above, however, he makes this claim regarding compulsory moral bioenhancement, and is not, in theory, opposed to voluntary moral bioenhancement.

[56] Nozick asks us to imagine being able to connect ourselves to a hypothetical machine that would simulate a life, designed and chosen to our personal specifications. Once connected, we would not know that the life we were living is a product of the machine, as it would appear entirely genuine. After a specified length of time, we would be able to disconnect from the machine and change particulars to satisfy our personal desires. The most important thing in the thought experiment is that the experience machine would be designed to maximise our happiness and would therefore be congruent with utilitarian precepts.

likely that most individuals would not endorse the God machine. The intuition that experience is only authentic and truly meaningful if rooted in reality and genuine choice and action, is analogous to the intuition that morality is only truly moral if it is unencumbered by external or artificial interventions. However, intuitions cannot, of course, be simply accepted; they require deeper investigation.

In his book, Harris also investigates the God machine hypothesis. He firstly discusses Savulescu and Persson's description of a similar thought experiment devised by Frankfurt (1969). This hypothesis entails an individual deliberating upon whether or not to do the morally correct thing by weighing up pertinent reasons in support of both options. After having come up with conclusive reasons in support of the morally preferable course of action, he then performs the morally preferable action – or refrains from performing the harmful action. It just so happens that if he had decided to commit the harm, what Savulescu and Persson describe as a *freaky mechanism*, based on Frankfurt's thought experiment, would have been activated in his brain to prevent this (2012:114). Savulescu and Persson argue that if the individual were to have chosen the morally correct course of action, thereby not requiring the freaky mechanism to intervene, his choice would hold equal moral value to a choice made in the absence of the presence of the freaky mechanism (ibid.). In other words – and to use Milton's phrase – despite the individual not being "free to fall" (ibid.), if he chose the morally correct course of action, there is no moral difference between his having done so in a context in which it would have been impossible not to do so, and one in which he had the option of 'falling'.

As mentioned above, Savulescu and Persson have based both their God machine and freaky mechanism thought experiments on the work of Frankfurt. Frankfurt discusses the common intuition, that, in cases in which an individual lacks the capability to choose between different actions, there can be no moral responsibility (1969:829). More specifically, he investigates the "principle of alternate possibilities" (ibid.) which outlines the claim that "a person is morally responsible for what he has done only if he could have done otherwise" (Frankfurt, 1969:829). Frankfurt argues that while the principle of alternative possibilities is generally accepted in philosophy as a form of "a priori truth" (1969:829), it is in fact, false. This is because he suggests that there could be contexts in which we would agree that an individual would bear moral responsibility for her actions, regardless of whether or not she had different choices available to her (Frankfurt, 1969:829-830). In order to illustrate his argument, Frankfurt looks at the phenomenon of coercion (1969:830). In contexts in which an individual is coerced to act in a

particular manner, we would generally agree that her freedom has been compromised in some way, thus impacting her moral responsibility regarding the action, and any outcomes produced by it. Frankfurt however, contests this intuition with an interesting thought experiment.

Imagine an individual, person A, who must choose between two courses of action, action X and action Y. Another individual, person B, strongly wishes that A choose to perform action Y; so much so, that he has decided to threaten A's family with death, should A refuse. However, B has decided to wait until A has made her decision. Only if A happens to choose action X, will B reveal the terrible threat and thus use it to force A to perform action Y. It just so happens, that A decides to choose action Y, thereby removing the necessity for B to reveal and exert his threat. Whilst this is an instance in which there is a clear existence of a coercive threat, and thus, A does not have a true choice of action available to her, Frankfurt argues that it seems obvious that we would hold A morally responsible for her decision to perform action Y. This is because A believed, even if falsely, that she had different possible actions available to her and *freely* decided to choose action Y. In other words when she made her decision, A believed that she had the possibility to have acted otherwise. Furthermore, as Frankfurt argues, if any praise was awarded as a result of the positive consequences of A performing action Y, it would be reasonable to award this to her, regardless of whether or not she genuinely could have acted otherwise. In other words, if we found out afterwards, that A did not truly have a choice, as, if she had chosen action X, the revealed threat was such, that she would have most definitely changed her choice to option Y, this wouldn't cause us to withdraw our esteem of her actions. In this way, Frankfurt does not view A's moral responsibility as having been eradicated purely due to the fact that she did not truly have a choice of action available to her.

The contested point here, however, is the notion of freedom. Person A may have believed that she was free to choose either action X or action Y, but in actual fact, she was not. It was the mere coincidence of her decision that was able to uphold the illusion of freedom. Simkulet makes a similar point when he engages with a slightly different version of this thought experiment. It is useful to apply his insights to the above example. In the above example, it is not the case that person A has no choices or alternative possibilities available to her, as Frankfurt implies. A can choose to perform action X or action Y. What A lacks is "alternate outcomes" (Simkulet, 2012:17). If A chooses action X, then B will intervene and prevent A from doing so, due to the nature of the threat. Thus, the only possible outcome is action Y; however, in the absence of this knowledge, A is making a genuine choice and this is why we would hold A morally responsible for this choice if

we were to know that it was made in the absence of knowledge about the impending threat from B.

Harris also wholeheartedly disagrees with Savulescu and Persson and Frankfurt's position on this matter. In order to elucidate his disagreement, he discusses Locke's example of *the locked house* (Locke, 1975; Harris, 2016:93). He asks us to imagine being locked inside a house, and thus, not being free to exit. However, what if you were inside this locked house and did not know that you were locked in? Furthermore – as implausible as this may sound – what if you never desired to leave the house? The salient question here is whether or not you would be free. It seems self-evident that you would not be free, regardless of whether or not you wished to ever leave the house. In the same way, Harris' claim implies that person A's decision, in Frankfurt's example, to choose option Y, was not truly a free decision. Rather, it was merely a coincidence that A happened to choose option Y. Here, Harris points out that there is a difference "between being capable and not being thwarted or frustrated" (2016:93). A person who is in a locked house and never decides to leave will not be thwarted in their actions in any way. However, at the same time, he will not be capable of leaving. Similarly, a person who has a freaky mechanism in their brain or is hooked up to the God machine and never forms the desire to commit harm will not be "thwarted or frustrated" (ibid) in their actions. However, regardless of whether or not they ever form the desire to harm, they lack the capability to act in this way, and thus, there is an obvious sense in which they are lacking in freedom. In other words, an individual who is not thwarted or frustrated in her choice to perform an action, but lacks awareness of important information such as the fact that if she happens to choose to perform a different action, she will be prevented from doing so, is not truly free.

In addition, by emphasising their point that an individual who lacks the capability to do wrong is as laudable for doing the morally good thing, as someone who could have done wrong and yet chooses to do right, Harris posits that Savulescu and Persson are missing the point regarding what actually matters about morality (2016:94)[57]. For Harris, what is important about morality is not

[57] Walker, an enthusiastic supporter of moral bioenhancement, makes a similar point that can be used in support of Harris' point, although he would not intend to do so I am sure! To those who would argue that moral bioenhancement would impact upon morality, because morally good acts would be easy to perform, and morality requires effort, Walker responds by arguing that even if we could establish that we do regard individuals who have acted morally, despite temptations, as more praiseworthy than those for whom acting morally is 'easier', this would never result in us deciding against inculcating moral codes and virtuous behaviour in our children (2009:39). In other words, the fear that our children would be less morally praiseworthy as a result of having internalised a sense of right and wrong, and thus, that they would be more likely to enact this in a relatively automatic manner, would not cause us to abstain from educating them in this manner. The point is that morality is not predicated upon praiseworthiness. The latter is

"praiseworthiness" (2016:94) for having performed good moral actions, it is freedom or agency. Furthermore, it is not sufficient to claim that moral agency remains intact so long as an individual is able to deliberate and have good reasons for his morally relevant actions or decisions. If we separate the process of moral deliberation, or reasoning, that informs our decisions to act in a particular manner from the actual decisions or actions themselves, in order to ensure that no harmful decisions or actions can be performed, then the process of deliberation becomes merely hypothetical, and thus, meaningless. As Harris has argued, freedom lies in the space between deliberation and decision-making and the action that it informs or leads to. Furthermore, as Harris points out, Savulescu and Persson may have a case for their claim that the capability to do wrong is not vital for moral deliberation to be free, however, it is not true that it is not vital "for moral choice *and action*" (ibid.) to be free.

However, Persson and Savulescu disagree with Harris in this regard, and they illustrate this by providing their own interpretation of the status of an individual in a locked house. They argue that if you deliberated about leaving and then decided to act upon your deliberations and leave, only to discover that you could not leave, then you would, of course, not be free to perform the act of leaving (Persson & Savulescu, 2016:265). However, if you deliberated about leaving – believing that you could leave – and then decided that your ensuing action would be to remain in the house, then both your decision and your act of staying would be freely made (ibid.). In fact, they posit that your freedom to stay, in the case of the locked house, would be as free as if you had decided to stay in a house that was unlocked. They therefore disagree with Harris regarding the fact that freedom requires not only freedom of deliberation, but also, the presence of alternative possibilities regarding action. Their argument here seems counter-intuitive, however. It seems that what Harris is getting at is that freedom must bear some genuine relation to reality. If an individual in a locked house, who does not know that he is locked in, then decides to stay in the house, believing that he is free, he is simply mistaken. His beliefs regarding reality, and thus, the fact that he is free to go, are false.

Ultimately, however, as Harris points out, Savulescu and Persson's God machine hypothesis reveals a fundamental clash in values (2016:106). As mentioned above, it is possible to defend the God machine hypothesis, and essentially, softer forms of moral bioenhancement, by arguing that they do not threaten moral autonomy. Or, one can take an alternative approach of defence, and

---

something we may bestow on moral acts when evaluating them, but it has nothing to do with the nature of morality itself.

argue that if moral bioenhancement did threaten our moral autonomy, this may not be a bad thing, all things considered. In other words, there may be certain things that are worth trading a portion of our moral autonomy for. Savulescu and Persson hold such a view and argue that "the value of human well-being and respect for the most basic rights outweighs the value of autonomy" (2012:411). Harris, however, disagrees with this position, arguing that "autonomy is a basic right quite as much as is freedom from violence or certain levels of well-being. Indeed, for him, autonomy (not just the illusion of autonomy) is part of well-being" (2016:107).

4.8.5 Liberty as Licence versus Liberty as Independence

Regarding Persson and Savulescu's use of Mill to support their views concerning the ultimate importance of well-being over autonomy, Harris argues that they are conflating two different types of liberty and thus failing to recognize the form that is truly important, and at stake, in the case of moral bioenhancement. Savulescu and Persson have argued, in various publications, that we are inured to impacts on our freedom due to various societal laws that regulate our behaviour, with various threats, including the loss of freedom, for any infringements (2012:411; 2014a:251; 2015b54; 2016:267). This distinction is important as it is used to provide impetus to their claim that the God machine's – and other forms of moral bioenhancement – prevention of serious harms, by making them impossible, would not be that different from other accepted mechanisms that aim to deter harmful behaviour. Thus, the argument would be that if we accept traditional restrictions on our freedom, we should also accept restrictions on our freedom caused by moral bioenhancement.

Harris disagrees with these claims, however, and discusses Dworkin's distinction between *Liberty as Licence* and *Liberty as Independence* to illustrate his point (Dworkin, 1977:262). Liberty as licence refers to the ways in which individual freedom is constrained by societal laws and restrictions. Without doubt, such laws do impinge upon liberty, however this impact is generally permitted due to the fact that it protects or supports "some competing value, like equality, or safety, or public amenity" (Dworkin, 1977:262). Liberty as independence, on the other hand, is a more fundamental kind of freedom. Dworkin describes it as referring to "the status of a person as independent and equal rather than subservient" (1977:262.). In other words, respect for this more existential type of liberty, would entail recognising the right of individuals to decide how best to live their lives, and to choose what they hold to be of value, including which morals, values, ideals and principles they espouse. As Dworkin points out, the two kinds of freedom may sometimes coincide. For example, onerous restrictions on liberty as licence may be employed to marginalise

or disempower certain groups within society, which may then impact upon their liberty as independence (1977:262.). Dworkin however, sees the two types of liberty as different in kind.

Persson and Savulescu seem to prevaricate regarding the particular conception of liberty they are addressing in their responses to Harris and other opponents of moral bioenhancement. With their argument that moral bioenhancement, and even the God machine, would be no different in kind to the societal restrictions on our liberty that we readily accept, Persson and Savulescu are clearly engaging with liberty as licence. That this is so, is clear in one of their papers where they explicitly argue that while the God machine would impinge upon our "agency and freedom, it would do so to a lesser extent than does the current penitentiary system, with such measures as imprisonment" (Persson & Savulescu, 2015b:54). However, elsewhere Persson and Savulescu concede that the God machine would "go further than law and its enforcement: it removes not merely the freedom to perform gravely immoral acts, but even the ability to do so" (2016:267). Where we have the option of breaking laws, and must therefore accept the relevant punishment if caught, in the case of the God machine, this option is non-existent.

In addition, in another publication with Douglas, Persson and Savulescu concede that there are different ways, and levels, in which autonomy may be constrained. One can constrain autonomy "externally (say through incarceration)…[or] internally" (Savulescu, Douglas & Persson, 2014:107), as would be the case with the manipulation of brain states via moral bioenhancement. Imprisonment prevents us freely moving around in the world, but moral bioenhancement may alter "bodily (brain) and mental states…[and in this regard] we have a stronger claim to bodily and mental non-interference than we do to freedom of movement. Thus…[moral bioenhancement] might seem to be a more serious restriction on autonomy than incarceration" (Savulescu, Douglas & Persson, 2014:107). However, while they make this concession, they argue that it is not self-evident that the differentiation between "external and internal" (Savulescu, Douglas & Persson, 2014:107) constraint has ethical relevance due to the fact that long term restrictions on movement, as experienced with imprisonment, do impact upon inner, mental states (ibid.). In addition, in the case of criminal offenders, constraints on internal autonomy may be justified in terms of the utility to society arising from the avoidance of serious harms that arise due to criminal acts.

Returning to Dworkin's distinction between liberty as licence and liberty as independence, for Harris, it is clearly the deeper, more existential, liberty as independence that would be threatened by moral bioenhancement, as Persson and Savulescu present it (Harris, 2016:108). Furthermore,

he argues that Persson and Savulescu's use of Mill's harm principle to support their view that well-being – where well-being is synonymous with harm prevention which moral bioenhancement would supposedly aim to support – is unequivocally more fundamental in value than autonomy, is a misconstrual of Mill's thought on the matter (Harris, 2016:107). Here, Harris draws support from Dworkins' reading of Mill. Mill, as discussed above, argues that "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others" (Mill, 1863:23). Persson and Savulescu take this to imply that the prevention of harm supersedes all other endeavours. However, in stating this, Dworkin posits that Mill is not advocating that all other forms of the exercising of power in society are illegitimate. In other words, he is not arguing that we do away with all restrictions on liberty as licence. This is because, along with protection from harm, liberty as licence may also serve to protect liberty as independence. Dworkin argues that Mill's harm principle is in fact concerned with liberty as independence and with the recognition of the equality of all individuals (1977:263). Dworkin argues that for Mill:

> an individual's independence is threatened, not simply by a political process that denies him equal voice, but by political decisions that deny him equal respect. Laws that recognize and protect common interests, like laws against violence and monopoly, offer no insult to class or individual; but laws that constrain one man, on the sole ground that he is incompetent to decide what is right for himself, are profoundly insulting to him. They make him intellectually and morally subservient to the conformists who form the majority, and deny him the independence to which he is entitled. Mill insisted on the political importance of these moral concepts of dignity, personality, and insult. It was these complex ideas, not the simpler idea of licence, that he tried to make available for political theory, and to use as the basic vocabulary of liberalism (1977:263).

With this interpretation in mind, Harris posits that Persson and Savulescu's claim that "the paternalism" (2016:107) of the God machine would receive justification from a reading of Mill, is, in fact, incorrect.

Regarding the God machine hypothesis, however, as I have stated above, Savulescu and Persson are not proposing that such a machine would ever be a real possibility (2016:267). Rather, they use it as an *intuition pump* in order to elucidate "whether there is anything that could rationally justify the negative gut reaction most of us are prone to have to it" (Persson and Savulescu, 2016:267-268). It seems that in the case of the God machine, the threat to moral autonomy is blatantly evident and cannot be argued away. This is perhaps why Persson and Savulescu focus on justifying their claim that autonomy should not be regarded as an absolute or fundamental value.

However, the God machine example also illustrates the point that restrictions to moral autonomy occur on a spectrum, with indeterminism – or absolute freedom from constraint – at one end, and determinism – or absolute constraint on freedom – at the other[58]. A compatibilist interpretation of freedom would regard our moral autonomy – as the product of traditional moral enhancement which includes internalised norms, cultural codes and general socialization – as lying somewhere in the middle of the spectrum. Such an interpretation acknowledges the power of causal influence – both physiological and social – on our actions, but sees these influences as compatible with our having freedom, or free will, to make choices. Being hooked up to the God machine, would, in all likelihood, move the status of our moral autonomy close to the determinist end of the spectrum. As mentioned above, Persson and Savulescu do freely admit that the God machine would not constitute a form of moral bioenhancement as it would "not enhance your motivation to do what is morally right" (2016:265). Rather, through entirely removing the ability to perform certain harms, it would be a form of behaviour control. The task remains however, to investigate exactly where along this spectrum of moral autonomy the softer forms of moral bioenhancement that are discussed in the literature, would lie.

Persson and Savulescu's various discussions of the God machine are interesting, however, as they make explicit their rejection of the truth of the *principle of alternative possibilities* (2016:265). This principle, coined by Frankfurt, and mentioned in section 4.8.4, posits that "moral obligation, responsibility and freedom presupposes that you can act otherwise than you in fact do" (Persson & Savulescu, 2016:265); or, as formulated by Frankfurt, "that a person is morally responsible for what he has done only if he could have done otherwise" (1969:829). Frankfurt, as discussed above, has argued for the falsity of the principle of alternative possibilities by providing examples of situations in which a choice rendered with no alternative possibilities could still be a choice that one could be held morally responsible for. The implications of this for our conception of freedom, if true, would be major. If it could be established that we hold individuals morally responsible for their actions despite their having no choices in certain situations, then it would imply that freedom does not solely depend on having alternative choices available. This, in turn, would have major implications for moral bioenhancement, and perhaps even the God machine. Frankfurt's example was not entirely convincing however, as was illustrated by Simkulet and Harris' responses.

[58] See footnote 60 for an explanation of determinism, indeterminism, compatibilism and incompatibilism.

4.8.6 Subjective versus objective alternatives

Due to the speculative nature of the subject, Persson and Savulescu make extensive use of thought experiments and analogies that serve as intuition pumps in their arguments and responses to Harris and other thinkers concerned with the threat to moral autonomy posed by moral bioenhancement. They have, however, provided a "theoretical backing" (Savulescu, Douglas & Persson, 2014:105) to their claims regarding autonomy, that is based upon a Kantian account of autonomy. They firstly point out that there is more to the notion of autonomy than simply referring to being able to choose between options (ibid.). While autonomy is generally understood as referring to the ability to steer or govern one's own life, they argue that autonomy is not simply "intentional" (Savulescu, Douglas & Persson, 2014:105). It also has an "evaluative…[and] normative" (ibid.) component. In order to fulfil this latter component, an individual must have knowledge of the possible choices available to her, as well as any information that is relevant to these different choices. Furthermore, autonomy requires that an individual must be able to "rational[ly] deliberat[e and] form rational beliefs" (ibid.). By using the term 'rational beliefs', they mean that autonomy presupposes that no "errors of logic" (ibid.) have been made and that the individual has a correct grasp of all accurate and relevant information. Furthermore, an individual must utilise her imaginative faculties in order to consider all the possibilities and outcomes associated with different available options. In addition, rational beliefs are the source, and impetus, of rational preferences and desires. This interpretation implies that preventing an individual from performing an immoral act that arises from an "irrational" or aberrant desire would not impact upon her autonomy (Savulescu, Douglas & Persson, 2014:105).

While having a choice implies having at least two options to choose from, Savulescu, Douglas and Persson distinguish between having "subjectively and objectively available alternatives" (2014:106). This distinction is particularly relevant for the position they take regarding the locked house example. Subjective choices refer to the choices that the individual regards as being available to him. Objective choices are the *actual* choices available to him. The two do not always coincide as an individual may regard himself as having no subjective choices in a particular situation in which he possesses objective choices. To illustrate with an example; I could find myself in a situation in which I have a choice between protecting myself from a terrible harm, or passing the threat onto a loved one to save myself. In this situation, I may have two objective choices available – to be harmed myself or to let my loved one be harmed – but no available subjective choices, due to my unwillingness to inflict harm on a loved one. Savulescu, Douglas and Persson argue that individuals are only autonomous, in the sense of self-governing, if they have

subjective choices available, from which they can then decide which option is best-suited to them (2014:106). Therefore, it isn't sufficient to posit that autonomy is present simply because an individual has objective alternatives available to him. In the case of the locked house, if an individual falsely believes that he is in a locked house, he would have no subjective options available, other than passively staying; whereas he would have two objective options available, to leave or stay. Conversely, an individual who chooses to stay in a house that is locked, not knowing that this is the case, will have two subjective options available, but only one objective option available.

As mentioned above, a Kantian account of autonomy requires that the individual possesses accurate information and beliefs, the absence of logical errors and the presence of imaginative faculties. If this is correct, however, then it undermines their argument in the preceding paragraph, as according to this conception of autonomy, an individual who has a belief that he is free to leave a house that is actually locked, has predicated his belief on false information. In other words, he may believe that he has two subjective choices available; however, this belief is false. It seems that for the distinction between subjective and objective choices to adequately describe the conditions for autonomy one would have to rather posit that individuals are autonomous when they possess *both* subjective and objective alternatives for action.

Savulescu, Douglas and Persson, however, argue that moral bioenhancement, if used in support of the above-mentioned requirements of autonomy, could possibly support, or enhance autonomy (2014:107). In other words, if a combination of moral education and moral bioenhancement is able to correct, or enhance, decision-making, then there is an argument that this will increase autonomy rather than compromise it. They posit that "even in cases of competent adults who have not consented to its use…[this] may not offend autonomy" (ibid.) if it produces improvements in the above faculties. It is possible to see how Savulescu, Douglas and Persson could come to the conclusion that enhancing the mechanisms associated with the conditions for autonomy would be supportive of autonomy, rather than inimical to it. It is, however, a long stretch for them to then argue that even in cases in which no consent has been given, that autonomy would not be impaired.

4.9 Three conditions for autonomy

DeGrazia, another noted supporter of moral bioenhancement, has also engaged with the potential threat to autonomy posed by moral bioenhancement (2014). He points out that it is seemingly the purported target of moral bioenhancement, suggested by its proponents, that Harris has a problem

with, rather than the idea of moral bioenhancement itself (DeGrazia, 2014:365). This target is, of course, the enhancement of motivation. As is evident from discussions in preceding chapters, morality is a notoriously complex and multifarious notion. It requires interplay between an abstract understanding of right and wrong and the ability to reason and infer from relevant contextual information, as well as the motivation or will to act upon this, thus resulting in moral behaviour. For Harris, what is pertinent in assessing morality is not motivation, but rather, the quality of moral judgements, the latter of which, for him, lie at the heart of morality.

However, in his response to DeGrazia, Harris argues that he does not lodge his critique of moral bioenhancement specifically against the enhancement of motivation. Rather, his argument is that moral bioenhancement, in the rudimentary and imperfect way in which it is currently possible, and the ways in which it is speculated to work in the future, would not "operate on anything so sophisticated and complex as 'motivation'" (Harris, 2014:372). Rather, the enhancement of so-called motivation will heighten emotion, making it more likely that an individual acts in a particular way due to the intensity of the felt emotion. This may, of course, happen to lead to morally desirable outcomes in some, or even most, situations; however, the resulting behaviour would not be moral behaviour. Harris argues that "one can do good accidentally but one cannot be moral accidentally" (2013:118). In other words, one may act in a way that has moral relevance in terms of the consequences it produces, but this does not make the action or behaviour "moral behaviour" (ibid.). For Harris, to qualify as moral behaviour, an act must be performed with moral intentions and this will only be the case if the behaviour is the product of a "moral judgement" (2014:372).

In terms of his concern that moral bioenhancement, or at least the administering of psychopharmaceuticals such as SSRIs or oxytocin, would impair our ability to make reasoned, "all things considered" (2014:372) moral judgements, Harris' concern is that such drugs clearly do produce behavioural changes and effects in those who take them. That this is so is evidenced by the fact that they are so widely prescribed for this express purpose. Of course, Harris is not lodging an argument against the correctly prescribed therapeutic usage of such drugs to treat psychopathologies; rather, his point is that calling them a moral bioenhancer is problematic and inaccurate. He points out that there is truth in the fact that:

> the presence of these molecules in particular doses is indeed 'freedom-subverting' if it were not, it is unlikely they would have the effects vaunted by their advocates, that is, effects that operate independently of the will or of judgement; unlike education, for example, that provides the basis for voluntary choices" (Harris, 2014:372)[59].

[59] Harris' argument that the administration of such 'molecules' is freedom-subverting is strange as it implies a bioconservative view that is not present in his work as an advocate of bioenhancement in general. In fact, the opposite

DeGrazia outlines his conception of "free action" (2014:366) as requiring three conditions. Firstly, an individual is only autonomous if her action is the product of having a clear preference to act in that way. Secondly, this preference must have been knowingly and thoughtfully assimilated by the individual. Thirdly, this preference must not arise, or be influenced by sources that she would reject, or take to be "alienating" (ibid.), after consideration. DeGrazia argues that it is not definitive that actions which are a product of freely chosen moral bioenhancement would fail to meet these conditions (2014:366). His delineation of the requirements for autonomy is seemingly sufficiently stringent, and in particular the third stipulation, so much so, that it is not obvious that many of us would be considered autonomous in our actions if the basis for truly free autonomous action is that our preferences must be free from alienating influences. In addition, such a requirement presupposes that individuals have a modicum of emotional insight and awareness of their behaviour and of the complexity of causes that exert an influence on it. However, while DeGrazia's conditions would require an accompanying theory or justification to substantiate them, the notion of autonomy as connected with having a preference to act in a particular way is one that I take to be of crucial importance and will therefore investigate in detail in chapter 5.

Harris also questions the third condition, arguing that it is characterised by potential "ambiguity and subjectivity" (2014:372). His concern is that an individual may not consider his preferences alienating (ibid.). The example Harris provides is of an individual with a racist disposition who may, in fact, see his beliefs regarding other races as justifiable and correct. In such a situation, the individual's racist beliefs may be a product of his upbringing and he may have internalised these beliefs as truths about the world, to such an extent that he is essentially unfree in his actions that arise from such false beliefs. Here, Harris is once again implicitly making the point that many immoral actions are the product of incorrect beliefs. Furthermore, Harris disagrees with another example provided by DeGrazia. DeGrazia discusses an individual who has been morally bioenhanced and who then acts according to the dictates of morality by providing assistance to someone in need, despite the fact that it would be less troublesome to refrain from providing such assistance (2014:366). In this situation, he posits that after consideration of her reasons for doing

---

claim could be made. For an individual suffering from a condition such as generalised anxiety, a psychopharmaceutical that reduces his anxiety levels would possibly enable him to deliberate more rationally and objectively. This is because the subjective experience of high levels of anxiety is akin to the experience of a strong emotion of fear, worry or general unease. Constantly experiencing such feelings would, most likely, compromise an individual's ability to correctly assess different situations and phenomena. This is a similar argument to the one Douglas makes when he posits that the reduction of counter-moral emotions would increase, rather than diminish, our freedom. It must be remembered however, that we are extrapolating from the effects produced by the therapeutic usage of such drugs. Using psychopharmaceuticals for enhancement purposes would be a different matter.

so, she would possibly recognise that her preferences have changed, that she has become more altruistic. She would possibly also realise that this change was due to the moral bioenhancement she had chosen to take. In this way, DeGrazia posits that there is no reason to assume that this individual would then consider this source alienating, thereby rendering her unfree (2014:366).

Harris disagrees with DeGrazia's claims here, regarding them as "disingenuous" (2014:372). He points out that the individual may well recognise her moral bioenhancement as the source of her changed desire to offer assistance and she may accept it and not regard it as alienating. However, her acceptance could very well be as a result of "the influence of the influence, not because it is the right thing to do" (ibid.). In other words, we, and the individual in question, would have no way of knowing whether her consideration and judgement of whether or not the influence of her changed behaviour is alienating, would be a true and free consideration. This consideration and acceptance of the source of her changed behaviour could be entirely a product of the intervention. Targeting motivation, in the way outlined by the proponents of moral bioenhancement, amounts to changing "attitudes" or preferences, in a way that individuals are more likely to act in accordance with this changed attitude or preference. In this way, for Harris, manipulating "motivation[s] does not meet standards of moral reasoning for the simple and sufficient reason that it does not meet standards of reasoning at all" (2014:372.). Furthermore, it is not only deliberation and moral reasoning that would be circumvented in such interventions, but morality itself, that would ultimately be removed from the process, he argues.

In terms of the definition of morality offered at the beginning of this section, DeGrazia concedes to Harris that if we are to define what makes moral behaviour moral, it is not only the presence of moral understanding and the ability to reason and infer from relevant contextual information, as well as the motivation or will to act upon this that are required; but also, freedom, or at least, "sufficient freedom" (2014:367) is a necessity. However, he only envisages a threat to freedom posed by certain exceptional cases of moral bioenhancement, and not in those interventions discussed in the literature. Furthermore, while DeGrazia concedes the point regarding the importance of freedom for morality, he agrees with the claims made by Persson and Savulescu regarding the need to balance the value of freedom against other goods, or at least the avoidance of harms. In support of this point, that he admittedly asserts rather than providing an argument for, DeGrazia suggests that while morality requires sufficient freedom to be truly moral, morality is valuable not only for intrinsic reasons but also for instrumental reasons (2014:367). In other words, we value moral behaviour not only as an end in itself, but also as a means of providing "a better

world with better lives for human beings and other sentient creatures" (DeGrazia, 2014:367). While intuition supports the truth of DeGrazia's assertion, Harris disputes whether morally bioenhancing individuals in the service of harm reduction would, in fact, constitute an improvement of the world and human lives (2014:373).

4.10 Other conceptions of freedom and autonomy

Bublitz argues that the way in which emotional enhancement is defined is crucial as it has direct implications for human autonomy (2016:89). Furthermore, there are also different ways in which freedom, itself, may be conceived. The advocates of moral bioenhancement engage with "freedom of action and freedom of will" (Bublitz, 2016:89), presenting arguments outlining why, and how, these freedoms will not be impacted by moral bioenhancement. Bublitz argues however, that there is a third conception of freedom, namely, "freedom of mind" which would be impacted by moral bioenhancement, and, which its supporters fail to engage with (ibid.). Freedom of action may also be understood, as Harris does, as *freedom to fall* or, at its most simplistic level, as the freedom to do wrong (2016:92). However, Bublitz agrees with the response of supporters of moral bioenhancement who frequently point out that this kind of freedom is limited due to the harm that may befall those who are affected by the actions of others who use this freedom for misdeeds. Thus, he argues, in a similar manner to Persson and Savulescu, that this conception of freedom has no normative force as it is accepted that "the freedom of one is inherently limited by the freedom of others" (Bublitz, 2016:92). It seems, therefore, that this conception of freedom is not what is at stake in the case of moral bioenhancement.

Beck also critiques the view that autonomy requires freedom of action (2015:237). In support of her claim, she draws upon Schmidt-Salomon's discussion, agreeing with his dismissal of the freedom of action due the fact that it is supported by a libertarian conception of freedom which requires metaphysical justification (2007)[60]. Schmidt-Salomon prefers a compatibilist

[60] A radical Libertarian conception of free will argues that our freedom is undetermined by any causal factors. In other words, such a view would support radical indeterminism, the view that human beings possess unfettered free will, due to the belief that choices, and thus actions and events, are not causally determined. Determinism, on the other hand, posits that all choices, and thus actions and events, are causally determined. A compatibilist is one who argues that free will exists despite the truth of determinism. In other words, such a view would argue that free will and determinism are compatible. Due to the fact that, generally, libertarians regard compatibilism as impossible, in other words they are of the view that determinism of any form invalidates our freedom, they would support indeterminism. However, there are a variety of differing libertarian interpretations of free will. In particular, a modern interpretation known as agent-causal libertarianism is perhaps what Schmidt-Salomon is referring to. This conception of free will would require metaphysical justification as it posits that an agent's actions are caused, not by any preceding or external events, but entirely, and inexplicably, by the agent in question. Schmidt-Salomon describes this conception of free will as a "hopelessly incoherent concept which should be dismissed for good reasons" (in Beck, 2015:237). The good reasons

interpretation and argues that the "freedom [that is] worth having is to be explained in terms of personal autonomy…[which refers to an] inner freedom of action (in Beck, 2015:237). This conception of freedom is similar to DeGrazia's conception of freedom in that Schmidt-Salomon describes it requiring "the absence of insurmountable internal compulsions with which the person is not able to identify positively" (ibid.). In other words, this conception of freedom is predicated upon having a preference to act in a particular way, and this preference must have been consciously formulated, rather than having arisen from coercive, compulsive or alienating sources.

While the first freedom that Bublitz discusses focuses specifically on action, freedom may also be conceived as entailing the ability to have and make *choices* between different possible actions, namely, the freedom of will (Bublitz, 2016:93). Faust has also discussed the freedom of will. She defines it as associated with being able to "independently assess circumstances (and all that make up circumstances), and autonomously choose action" (Faust, 2008:404). The part of this process that is not fully understood is how we move from our observations and deliberations to actually acting on them. Faust posits that this unknown area is akin to what Searle has called "the gaps in rationality" (2008:405). Searle has argued that rational acts are mostly characterised by a gap between motives and desires, and the decisions that they lead to. This gap would generally be associated with the "freedom of will" (Searle, 2001). As Faust points out, if we observe two individuals who are presented with identical facts regarding a situation, they will invariably not interpret these facts in the same way, and therefore, they will not necessarily act in the same way. However, as Searle posits, this is "exactly what makes us rational—what happens in the gaps between observation and assessment, assessment and judgment, judgment and action" (in Faust, 2008:405). Regardless of whether or not we understand the nature of this gap, we are able to impact it and this elicits concern regarding our free will. In the case of moral bioenhancement, it is not necessarily that an individual would be compelled to act in particular manner, rather, it could work akin to a reflex, or as Faust describes it, "an instinctual reaction to duck when an object is thrown at one's head" (2008:405). However, as she points out, despite our having certain reflexes, we are able to overcome them when we need to in specific contexts, for example, in certain sports like soccer when one is required to 'head' the ball rather than kicking it. If moral bioenhancement worked in a similar manner to such reflexes, then it wouldn't necessarily eradicate our free will.

---

he refers to here would presumably be the difficulty of explaining how such an agent entirely avoids the influences of causality, due to some mysterious internal component.

Returning to Bublitz' discussion of the freedom of will, he engages with Persson and Savulescu's arguments regarding the relevance of moral bioenhancement for the issue of determinism and indeterminism. Persson and Savulescu have argued that if one is a compatibilist, then moral bioenhancement will pose no threat to our autonomy (2013:128). In other words, if one believes that free will remains intact despite the constraints of causality – in this case being "determined whether or not we shall do what we take to be good" (Persson & Savulescu, 2013:128) – then our freedom would not be affected by moral bioenhancement. We would simply enhance ourselves to be more like those who generally always act in a moral manner. If, however, one supports the view of indeterminism, then one will most likely not regard moral bioenhancement as a means of improving the likelihood of choosing to do right more often. This is because the success of moral bioenhancement would be "limited by our freedom in this indeterministic sense" (ibid.). Thus, Persson and Savulescu conclude that on both accounts of freedom, moral bioenhancement will be exempt from impact.

Blackford has also addressed the above points made by Persson and Savulescu, arguing that to a certain extent, the concern that moral bioenhancement may undermine our autonomy will depend on whether one is a compatibilist or a non-compatibilist. In this regard, he argues that "we should be careful not to attribute to ourselves (and other ordinary people) a spooky kind of 'autonomy all the way down' that does not exist in the real world" (2010:83). In other words, we should not overestimate the extent to which we actually are autonomous, or possess free will, to begin with. Here, Blackford is seemingly supporting a determinist – and possibly an incompatibilist – account of freedom. To explicate his point, he discusses Galen Strawson's work regarding the issue of moral responsibility (1994). Strawson has argued that moral responsibility is an illusion due to the fact that the kind of person that we are, is a product of a complex entanglement of "our heredity and early experience…[as well as other] indeterministic or random factors" (in Blackford, 2010:83), the nature of which we have had no control over. Thus, all ensuing decisions taken when we reach maturity, including any attempts to transform ourselves, will be strongly influenced "by how we already are" (in Blackford, 2010:84). Therefore, if one agrees that autonomy implies "ultimate self-causation" (ibid.) and that the latter is an illusion, for the reasons provided by Strawson, then one will agree that autonomy is an illusion. However, a compatibilist would, of course, respond by disagreeing with this radical conception of autonomy, arguing that autonomy, and thus moral responsibility, does not require that we have absolute control over how we came to be who we are (ibid.). This latter view is the more prevalent one within contemporary free will debates in philosophy.

On the other hand, as Bublitz points out, empirical research in the field of neuroscience gives us grounds to also dismiss the possibility of an indeterministic conception of freedom. In terms of the compatibilist conception of freedom, however, it is not such a straightforward matter. There are other variants of compatibilism that are more nuanced than the interpretation presented by supporters of moral bioenhancement, such as Persson and Savulescu and DeGrazia (Bublitz, 2016:94). Here, Bublitz specifically alludes to DeGrazia's conception of freedom, discussed in section 4.9, to illustrate his point. DeGrazia argues that freedom requires that an individual's action be the product of having a clear preference to act in that way and this preference, in turn, must have been knowingly and thoughtfully assimilated by the individual, rather than having arisen from sources that the individual would reject, or take to be "alienating" (2014:366), after consideration. Thus, even if an individual's preferences were altered due to having undergone moral bioenhancement, this would not impact upon freedom if these changes were accepted and not regarded as alienating.

This may make *prima facie* sense, however, Bublitz points out that it fails to engage with an interpretation of compatibilism in which individuals would be exempt from moral responsibility if their preferences have been subject to "manipulation" (2016:94) in any way. Here, Bublitz cites the work of Fischer and Ravizza who have discussed examples in which ill-intentioned neurosurgeons have altered the preferences of individuals in order to cause them to perform morally dubious acts (1998:182 in Bublitz, 2016:94). In such cases, the individuals in question would not be held morally responsible for these acts and their consequences. The reason we wouldn't hold an individual who commits a harmful act morally responsible, upon acquiring knowledge that she had been neurobiologically manipulated in this way, is because we would regard her autonomy as having been impaired by the intervention. We would believe that she was compelled to act in this way, through no choice of her own, and that the intervention would have overwhelmed the possibility of her rationally considering her action. As argued by Fischer and Ravizza, autonomy cannot, thus, be viewed in an a-historical manner; the origins of how we come to have particular preferences are significant in assessing the preferences themselves, and the acts they lead to (1998:182 in Bublitz, 2016:94). Neurobiological manipulation would therefore be a clear example of a way in which autonomy would be impaired. Thus, there is a difference between a change in preferences due to a "direct brain intervention"[61] (Bublitz, 2016:94) and the examples that

[61] Here, Bublitz includes other mechanisms such as "hypnosis..[and] subliminal advertisement" (2016:94) as interventions that would thwart moral responsibility if they impacted upon individuals' preferences, and thus, their actions.

Savulescu and Persson give regarding women not being less free due to possessing more empathy. In such an example, "feminised men and women may have the same mental and moral properties, yet women are responsible, whereas manipulated men are not, not because of gender but because of their diverging histories" (Bublitz, 2016:94)[62]. This point is crucial for the line of argumentation that I will present in chapter 5 and will therefore be developed further.

While the afore-mentioned, conventional conceptions of freedom are engaged with by the supporters of moral bioenhancement, Bublitz argues that there is another, somewhat neglected, conception of freedom that holds great value for human beings. This freedom is one that may be threatened by moral bioenhancement (2016:94). Bublitz identifies it as the "freedom of mind, [which] is the freedom of a person to use her mental capacities as she pleases, free from external interferences and internal impediments" (ibid.). In comparison to freedom of will, freedom of mind includes a wider array of "mental states" (ibid.). As Bublitz points out, interpretations of free will are regarded as having relevance for stipulating the criteria under which individuals can be regarded as morally responsible for their actions. One of the criteria generally considered relevant in terms of ascribing moral responsibility is: could an individual have acted otherwise? This is the conception of moral responsibility supported by the principle of alternative possibilities that was mentioned in section 4.8.4, and which Frankfurt argues is false. Other important aspects of the will that must be present in order to be held morally responsible are: possessing sufficient cognitive capabilities in order to provide reasons for one's actions and not being impeded by "overwhelming inner constraints such as irresistible impulses" (Bublitz, 2016:95). Generally, if individuals can be shown to meet these criteria, we would regard them as morally responsible for the outcomes of their choices and actions. Bublitz posits that it would be possible to argue that this conception of freedom will not be threatened by moral bioenhancement.

However, freedom of mind differs from this conception in that it is not so much connected with the attribution of responsibility, but rather, it is concerned with attributes and capabilities in which individuals have a vested interest. The administration of psychopharmaceuticals, discussed by supporters of moral bioenhancement in the literature, would not necessarily have large scale, deleterious effects on individual freedom; rather, it is possible that they could "subtly alter perception, mood, emotional patterns or the style of thinking. They may subdue or increase the intensity of emotions, enhance or decrease mental skills or alter mental background conditions"

[62] This is a similar argument to that made by Harris, discussed in section 4.9, who argues that an individual who has been morally bioenhanced would not necessarily regard her preferences as alienating because her preferences would, themself, have been altered by moral bioenhancement. Thus, DeGrazia's conception of freedom is not sufficient.

(Bublitz, 2016:95). Thus, while what is at stake in the case of freedom of mind is not a decisive loss of autonomy, Bublitz posits that there are nevertheless ethical concerns that require investigation.

To provide more substance to his conception of freedom of mind, Bublitz discusses what he regards as its key components. The first component is "conscious control over one's mind" (Bublitz, 2016:95), referring to the extent to which one's mental capabilities are free from internally constraining elements, such as excessive emotional responses, which is, of course, the same concern that Harris has against moral bioenhancement. While Bublitz concedes that both emotional and cognitive responses are vital for moral judgements, he nevertheless argues that any intervention that intensifies emotional responses would be problematic if it overwhelms our ability to consciously control our minds (2016:96).

Focquaert and Schermer hold a similar view of autonomy, agreeing with Harris that moral autonomy requires that one's actions are motivated by reasons (2015a:142). Without "responsiveness to moral reasons…[an] intervention should be understood as a form of behaviour control" (Focquaert & Schermer, 2015a:143). Focquaert and Schermer's understanding of autonomy entails "leading one's life in accordance with one's own choices, that is, choices that are based on the values and goals we endorse after deliberation" (2015a:145). In the case of moral bioenhancement, and with this understanding, preserving one's autonomy would require that if one agrees to an intervention there must be the possibility of being able to change one's mind and opt out if one so wishes. This would presumably include being able to reverse the effects of the intervention if one so desires.

In accordance with their concern, discussed in section 4.2.5, the types of interventions that could risk impacting upon personal identity would pose the same risks for our autonomy. In particular, passive interventions could pose a distinct threat to autonomy, even if consent has been given prior to undergoing such interventions. If interventions produce the kinds of hidden changes in personality, discussed in section 4.2.5, then an individual who has been morally bioenhanced could be less likely to withdraw consent. If the changes in personality were such that they caused the individual to accept the intervention, whereas consent would have been withdrawn if the 'original' personality were intact, then this would be a subversion of autonomy.

Returning to the second component of freedom of mind that Bublitz identifies which is "peace of mind and mental integrity" (2016:97), this refers to the entirety of those aspects of our mental life that we are not consciously aware, or in control, of. As their functioning occurs unconsciously, and thus automatically, these aspects are notoriously ephemeral and difficult to define or identify precisely. Bublitz explains them, however, as the way in which "our conscious states such as thoughts and feelings are processed, prepared, triggered and realised by unconscious mechanisms on several levels, from single neurons to networks of brain areas to supposedly higher-level psychological operations" (2016:97). Here, he draws on some of the findings of Crockett's research which indicates that certain interventions, targeting "emotional propensities, behaviour dispositions and preferences" (ibid.) result in changed behaviour in the absence of any accompanying changes in "judgements or preferences" (ibid.; Crockett et al. 2008; Crockett et al. 2010a; Crockett et al. 2010c). In other words, despite the fact that the subjects act differently, these changes can only be explained as having been the product of "an alteration of unconscious dispositions" (Bublitz, 2016:97), as these subjects were not aware of any conscious, cognitive changes that could explain their changed behaviour.

Bublitz admits that even in the absence of any interventions, we do not have control over our unconscious mental states. However, interventions that produce the kinds of changes described above would possibly threaten something that we have an interest in protecting: our peace of mind. He understands the latter idea as encompassing a "negative sense of mental freedom: to remain untouched from interventions tampering with consciously uncontrollable mental elements" (Bublitz, 2016:97). Bublitz explains this concern with an analogy regarding the automatic nature of our physiological systems, such as our hormone levels. Despite the fact that we have no conscious control over these systems, if one of them was inadvertently altered, without our awareness or consent, this would generally be regarded as "an "illegitimate violation…or more precisely…[a violation of one's] right to bodily integrity" (ibid.). In the same way, interfering with one's "self-regulatory mental system" (ibid.) would be similarly problematic, and thus, interventions that produce such effects require further ethical investigation.

The third component of freedom of mind is one that resembles the preceding components, namely, "the freedom to hold beliefs and opinions without interference" (Bublitz, 2016:98). As Bublitz points out, however, it merits discussion in its own right due to the fact that it is a freedom that is afforded protection in human rights instruments such as the Universal Declaration of Human Rights

(1948)[63]. Bublitz posits that due to the fact that it is widely accepted that freedom of belief and opinion are of absolute value, we should be wary of any intervention that would result in changes to our beliefs and opinions, particularly if such changes are not the result of "argument and persuasion" (2016:98). Here, Bublitz draws attention to Persson and Savulescu's arguments that subjects undergoing moral bioenhancement would not have their cognitive capacities subverted as they would still be acting in a reason-responsive manner. What will have changed post-moral bioenhancement, according to Persson and Savulescu, is that such individuals would now "act for the same reasons as those of us who are most moral today do" (2013:129). It is, however, this very "replac[ement of] moral reasons, motives, insights, beliefs and opinions" (2016:98) that concerns Bublitz. For Bublitz, like Harris, altering or intensifying emotional responses would be an example of an intervention that would directly change the afore-mentioned qualities while bypassing acceptable and ethically justifiable mechanisms of change such as rational argumentation and persuasion (2016:99).

Bublitz does point out, however, that one has the right to freely consent to interventions that would impact on the above-mentioned freedoms and that this would not be self-evidently problematic, providing such consent was genuinely free from coercive forces. Examples of coercive forces would not only be the obvious contenders, such as compulsory, state-sanctioned moral bioenhancement, but also less obvious forms of coercion, such as providing incentives to opt for moral bioenhancement that are difficult to refuse. The value of Bublitz's discussion is that it illustrates the need to give further clarification to the notion of freedom and autonomy and to organise the many interpretations that have been discussed in this chapter. This is a task that will be addressed in the following chapter.

4.11 Concluding remarks

In this chapter, I have presented a comprehensive overview of the way in which the concern for the impact of moral bioenhancement on moral autonomy has been addressed in the literature. A synthesis of the discussions and insights of this chapter reveals the following important questions and issues that must be addressed. The first question is the issue of whether or not – and if so, to what extent – moral bioenhancement would, in fact, threaten our moral autonomy. The answer to this question depends upon a number of factors. Firstly, it is dependent upon how moral autonomy, itself, is defined. On some interpretations, moral bioenhancement could arguably increase our

[63] See article 18 and 19 of the UDHR.

autonomy, whilst on other conceptions it could decrease or eradicate it. Thus, it is vital that the type of autonomy that is at stake is clarified.

Secondly, it is dependent upon the type of intervention that would be employed. This issue is, to a certain extent, connected with the problem of what the target of moral bioenhancement would be, and thus, with the problem of moral content, discussed in chapter 2. The interventions that have been proposed in the literature are diverse, and range from those that are less invasive, such as pharmacological interventions, to those that are highly invasive, such as genetic or neurological interventions, like DBS. In addition, because the science that would enable moral bioenhancement is not yet, and may never be, possible, this area is largely speculative in nature. As mentioned in section 4.2.5, if we view freedom as occurring on a spectrum, then interventions that are passive – requiring little or no cognitive input from the individual – or, possibly, those that act directly upon the brain or mental states producing behaviour changes in the absence of cognitive reflection, would be the most problematic. If such interventions were to produce decisive behavioural changes or *excessively* heighten emotional responses, they could be more accurately described as forms of compulsion or behaviour control. Furthermore, other problematic interventions would be those that produce hidden identity or personality changes, those that change core beliefs, those that alter the mechanisms by which we are able to assess any changes, and those whose effects would be irreversible, thus thwarting the ability to withdraw consent. In addition, interventions producing subtle changes, such as those associated with the freedom of mind could be problematic.

Thirdly, the answer is dependent upon how moral bioenhancement would be administered. While compulsory moral bioenhancement would be a clear violation of our autonomy, it isn't necessarily the case that voluntary moral bioenhancement would leave our autonomy entirely intact. Moral bioenhancement producing the above-mentioned effects and lying at the far end of the spectrum of freedom could be problematic regardless of whether or not it was administered voluntarily.

The second question is whether or not it would be a decisively negative thing, all things considered, if moral bioenhancement were to impact autonomy. The answer to this question is also dependent upon other factors such as the value that one ascribes to the freedom to do wrong and to autonomy in general. One's opinion regarding the latter will be influenced by how one ascribes value to autonomy. If autonomy is regarded as intrinsically valuable, then any impacts upon it would be regarded as impermissible; whereas if it is valued primarily for instrumental reasons, then such impacts would be acceptable if they were associated with positive outcomes. Of course, this

distinction is an over-simplification of how we ascribe value, the two conceptions inform one another and it is likely that if we value autonomy, we do so for both instrumental and intrinsic reasons.

However, this point aside, regarding the question posed above, it is likely that one would only pursue this line of inquiry further if one regards moral autonomy as either entirely, or partially, instrumental in value. If this were the case, then whether or not one would view impacts upon our moral autonomy as justifiable would depend upon the nature of the good that would be produced by moral bioenhancement, where good would be commensurate with the avoidance of harm that moral bioenhancement would avert. In other words, impacts on autonomy would have to be justified by means of a utilitarian calculus that would have to indicate, as reliably as possible, why a trade-off between safety or well-being and autonomy would be a worthwhile justification for a programme of moral bioenhancement. Due to the speculative nature of this matter, I will not pursue it further, as I wish to rather focus on the first question posed above which I take to be more philosophically interesting. I will therefore address this concern in detail in chapter 5.

Chapter 5a – Autonomy as authentic self-determination

5.1 Introduction and overview

There are a variety of ways in which one may approach the task of elucidating the concern for autonomy voiced by the opponents of moral bioenhancement. One could approach the concern from a neuroscientific paradigm in order to argue that the issue is a moot one, as neuroscientific knowledge of the brain is such that we now have evidence that we do not, in fact, possess the kind of autonomy that we assume we do. The philosophical equivalent of this approach would be to argue for incompatibilism. This is the position which argues that, to the extent that our actions are causally determined by physiological antecedents, we cannot possess free will, and thus, autonomy, in the sense of our ability to fully determine our lives, is an illusion. I will not take either of these approaches; rather, I will assume a compatibilist interpretation of free will, in order to be able to engage coherently with the autonomy concerns that have been lodged in the moral bioenhancement literature. The opponents of moral bioenhancement who argue that it will threaten autonomy clearly assume a compatibilist interpretation. In other words, if they supported incompatibilism, they would not fear any threats to our autonomy posed by moral bioenhancement, as autonomy would not be something that we possessed to begin with. One cannot lose something that one does not have.

The autonomy concerns in the literature take the form of philosophical arguments. Therefore, I have chosen to engage with them in an immanent manner. In other words, I will approach the matter in the same manner, from within, and with the same criteria, as these arguments utilise, to try and ascertain if the concerns are valid. In order to do so, I will need to provide an account of what is at stake, namely, an account of autonomy itself. Furthermore, this account of autonomy must be one that is suitably rigorous, if it is to be taken seriously as a means of resolving the concern for moral autonomy posed by moral bioenhancement. There are certain accounts of autonomy that have been extremely influential, but are nevertheless associated with various problems. In fact, there are three specific problems that have plagued theories of autonomy. If a theory of autonomy could be shown to have the ability to overcome these problems, it could be of immense help in achieving the above aim.

In other words, to be able to coherently assess the concerns put forward by Harris and other thinkers, one must do so by way of choosing a particular account of autonomy. One must also be able to provide evidence of the legitimacy of this particular account of autonomy by way of some

mechanism or tool that can be used to assess it. A theory of autonomy that is able to pass the test of the three problems of autonomy would be one such potential candidate. More specifically, for the purposes of an assessment of moral bioenhancement, one of the three problems, in particular, would be most relevant. Any theory of autonomy that is to be accepted as a legitimate means of assessing the concern for moral autonomy posed by moral bioenhancement would have to be one that did not fall foul to *The problem of manipulation*. This is because the problem of manipulation bears many resemblances to the processes by which moral bioenhancement would work, as alleged by its opponents.

If a theory of autonomy fails the test of the problem of manipulation, what this means is that it is unable to clearly indicate why external interference or manipulation would be a compromise to autonomy. In other words, if such a theory provides specific conditions for an action to be considered autonomous, and we are able to show that these conditions would be met even when some form of external interference has taken place, then it would be said to have failed the test for the problem of manipulation. A theory that passes this test would be one in which the conditions for autonomy are formulated in such a way that they exclude the possibility of external interference or manipulation. Furthermore, excluding manipulation simply by fiat is not sufficient; a theory of autonomy that is to be taken seriously as a potential tool for resolving ethical disputes must be able to give clear grounds as to *why* such external interference would compromise autonomy. Therefore, any moral bioenhancement intervention that can be shown to pass the test for autonomy when assessed by means of a theory of autonomy that has itself been shown to overcome the problem of manipulation, could possibly be ethically permissible.

I will therefore begin this chapter with a brief discussion, in section 5.2, of the concept of autonomy itself. This will be followed by an investigation of hierarchical theories of autonomy in section 5.3, focusing specifically on Harry Frankfurt's theory of autonomy. Such an investigation is necessary, as, since its inception in the 1970s, virtually all contemporary accounts of autonomy have been informed by, or have reacted to, Frankfurt's account. In particular, the theory that I have selected as the most effective means of addressing the concern for the threat posed by moral bioenhancement to autonomy – the coherence theory of autonomy – is founded upon insights original to Frankfurt's theory, while managing to avoid the pitfalls that have plagued his account. Thus, it is only possible to adequately grasp the coherence theory that I will utilise, if one possesses a full understanding of its predecessors. Furthermore, the strength of this coherence theory of autonomy, and its legitimacy in being able to assess the concern for autonomy posed by moral

bioenhancement, will only be evident if one understands just how challenging it is for a theory to overcome the three problems. In section 5.4 I will therefore discuss the three problems associated with hierarchical theories of autonomy, and the way that these problems have compelled thinkers to both adapt these theories, and to come up with new theories of autonomy. This will be followed by a discussion of Gerald Dworkin's "criteria for a satisfactory theory of autonomy" (1988) in section 5.5, which are a useful means of assessing and justifying a potential theory of autonomy.

In section 5.6 I will then introduce Laura Ekstrom's coherence theory of autonomy (1993, 1999, 2005a, 2005b). This theory, not only avoids the problems that plague earlier hierarchical theories of autonomy, but most importantly, it avoids the problem of manipulation. Furthermore, it is a theory of autonomy that is predicated on the assumption that one's personal or psychological identity is a determining factor in autonomy. Therefore, it illustrates the argument that was discussed in chapter 4a, namely, the concern that major impacts or changes to core aspects of personal identity could be problematic for autonomy. Once I have presented and discussed Ekstrom's theory, I will then assess it in general, as well as by means of Dworkin's criteria for an adequate account of autonomy in section 5.7. In the second part of this chapter I will then use the above insights to analyse the way in which different moral bioenhancement interventions could impact this particular conception of autonomy.

5.2 A brief discussion of autonomy

The concept of autonomy is wide enough in scope to be the subject of an entire volume. This is evidenced by the fact that the since the publication of several papers seeking to investigate the concept on a deeper level in the 1970s, there have, in fact, been several volumes specifically dedicated to the subject, which have had great subsequent influence. In this section, however; I will provide only the briefest discussion of some background information that I regard as relevant for understanding the concept in general, as well as salient for the argument that I will make in the second part of this chapter.

In the moral bioenhancement literature, there is very little in-depth engagement with the notion of autonomy itself. Most of the arguments that discuss the concern for moral autonomy are conducted at a somewhat superficial level, due to the fact that they proceed with their investigations after implicitly assuming consensus regarding the meaning of autonomy as something along the lines of moral self-determination in a Kantian sense. This tends to frame these arguments in such a way that the debate has become somewhat stultified. It is, of course, understandable that there is a

tendency to think that what is at stake in the moral bioenhancement debate is Kantian moral autonomy. This is because the argument is so frequently framed as the concern that moral bioenhancement, by way of intensifying our emotive responses, and thus, increasing our motivation to act in a particular manner, deemed moral, will cause morality to become infected by contingent factors rather than being predicated upon the dictates of the rational will. This outcome would be the antithesis of the Kantian notion of what it means to act morally. However, while it has enjoyed enduring influence, this Kantian conception is but one interpretation of moral autonomy.

Moral autonomy may also be framed in terms of *moral authenticity* (Feinberg, 1989:36). This interpretation is closely related to what I take to be at stake in the moral bioenhancement debate. Such an interpretation argues that the autonomous individual is not only one who is moved by his *own* general attitudinal states, but also one whose "moral convictions and principles (if he has any) are genuinely his own, rooted in his own character, and not merely inherited" (Feinberg, 1989:36). Using the requirement that value systems must be the individual's 'own', does not imply that we choose, or are *able* to choose, our moral value system *ex nihilio*, as this would be an impossibility due to the fact that our value systems are deeply influenced by the context in which we reach moral maturity. Furthermore, while this interpretation is similar to the Kantian interpretation, in terms of morality requiring that we act from autonomous rather than heteronomous reasons, it doesn't give substantive requirements regarding how this must be done, simply, that to be considered autonomous, an action may not be mindlessly performed or be caused by undue external impacts. Thus, what is meant by this interpretation is that a morally authentic person is one who is aware of, and has examined her attitudinal states, including her value system, in the light of rationality, and who makes any changes on the basis of good reasons that are hers (Feinberg, 1989;32). The notion of rational examination is therefore a thread that runs through most accounts of morality and is also regarded as a necessary component for the possibility of being able to exercise both morality and autonomy. However, an account of moral autonomy as authenticity differs from a Kantian account in that acting from one's own moral convictions and principles would include the role that emotional and attitudinal states play in this process.

Historically, the etymology of the concept of autonomy, *auto-nomos,* is Greek in derivation and its meaning can be traced back to the ancient context in which it signified the independence or self-governance of Greek city-states. The connation of the concept has not changed fundamentally since this initial use, as it is still understood to imply self-governance or self-determination. Rather, it is the denotation of the concept that has changed, as it now refers to the autonomy of *individual*

human beings rather than city-states. Autonomy is frequently used interchangeably with notions that belong to the same conceptual family such as freedom and liberty. However, while these concepts are related, they are characterised by subtle differences. As Christman succinctly formulates it, autonomy at its most fundamental "level of application…is more properly seen as a property of preference or desire formation than a property of whole persons or of person's whole lives" (1989:13). Here, the notion of preference can be taken to signify a variety of mental states such as attitudes of value and beliefs. This is contrasted with the notions of freedom or liberty which are "a property of human action – a characteristic of the relation among desires, bodily movements, and restraints that may be facing the agent" (ibid.).

Christman provides a formula for freedom, arguing that being "free (in a given context) means there is an absence of restraints (positive or negative, internal or external) standing between a person and the carrying out of that person's autonomous desires" (1989:13). This formula also clearly indicates the closeness of the relationship between freedom and autonomy. However, they are related in a complex manner as one may be free but not autonomous, and one may, possibly, be autonomous but not free. An example of the former situation would be an individual who possesses freedom of action, but, whose preference to live his life in a particular manner is thwarted by a strong addiction to use drugs. Examples of the latter are more challenging to find; however, one such example was discussed in the previous chapter in section 4.8.1: at the moment that he was tied to the mast, Odysseus was not free to move closer to the sirens as his will desired in that particular moment. However, he was autonomous with respect to his original desire and request to be bound to the mast and ignored should he ask to be freed. Another example would be an individual who is the slave of a benevolent master and has limits on his freedom of action, but is autonomous regarding his ability to form and act on most preferences. However, such examples are open to contestation, and whether or not an individual in such a situation would be regarded as possessing autonomy would depend entirely upon the particular theory of autonomy that is being utilised as a means of analysis. What is clear, however; is that autonomy is a more demanding phenomenon than freedom and it is a decidedly internal state.

Another useful interpretation that must be mentioned, due to its relevance in pinpointing the meaning of autonomy, is Isaiah Berlin's distinction between positive and negative liberty (1969). The notions of liberty and freedom are generally regarded as synonymous, and Berlin explicitly states that he regards them as interchangeable (1969:121). In its most simplistic formulation, negative liberty refers to freedom *from* constraints whereas positive liberty refers to the freedom *to*

do a particular thing. Another way of formulating the difference between the two is to conceive of negative liberty in relational terms, as it pertains to the way in which freedom may be impacted upon in the *interpersonal* realm, whereas positive liberty is an *intrapersonal* notion as it is concerned with the extent to which the individual is able to determine her life's course. However, these interpretations are over-simplifications to a certain extent, as the internal freedom to act in a particular way is affected by the extent to which one is subject to external constraints.

While Berlin remarks on the fact that the two notions are seemingly similar *prima facie*, he points out that their historical development was tangential and at times directly at odds with one another. Regarding negative liberty, Berlin posits that:

> I am normally said to be free to the degree to which no man or body of men interferes with my activity…If I am prevented by others from doing what I could otherwise do, I am to that degree unfree; and if this area is contracted by other men beyond a certain minimum, I can be described as being coerced or, it may be, enslaved (1969:122).

By including the phrase "what I could otherwise do" (ibid.), it is obvious that Berlin is not speaking of absolute freedom, as there are things that one simply cannot do, due to the fact that we are limited by our materiality. As a human being, I cannot fly with only the aid of my arms; I cannot hear if I have been born deaf; and, I cannot perform certain mathematical equations in the absence of a certain level of mathematical acumen. The type of liberty that Berlin is therefore engaging with here is specifically that freedom that can be threatened through human interference.

Berlin defines positive liberty in terms of:

> the wish on the part of the individual to be his own master. I wish my life and decisions to depend on myself, not external forces of whatever kind. I wish to be the instrument of my own, not of other men's, act of will. I wish to be a subject, not an object; to be moved by reasons, by conscious purposes, which are my own, not by causes which affect me, as it were, from outside. I wish to be somebody, not nobody; a doer – deciding, not being decided for, self-directed and not acted upon by external nature or by other men as if I were a thing, or an animal, or a slave incapable of playing a human role, that is, of conceiving goals and policies of my own and realizing them…I wish, above all, to be conscious of myself as a thinking, willing, active being, bearing responsibility for my choices and able to explain them by references to my own ideas and purposes. *I feel free to the degree that I believe this to be true, and enslaved to the degree that I am made to realize that it is not* (own emphasis, 1969:131).

From the above quote, it is evident why Christman describes Berlin's conception of positive liberty as the "identical twin" (1989:3) of autonomy. Berlin's conception captures what is regarded as valuable about autonomy, and, what is therefore at stake when it is threatened. In addition, the last sentence of the quote is particularly important for the discussions of autonomy that will follow. This conception of autonomy is a deeply personal one as it is linked to the subject's perception of the extent to which she is free to be moved by her own desires, beliefs and attitudes.

With the above comments in mind, it is my contention that only if a richer theory of autonomy – one that contains specific criteria against which interventions may be assessed – is utilised, will the moral bioenhancement debate be able to move beyond the current impasse. Furthermore, a more rigorous approach has major implications for other areas in bioethics that utilise a thin interpretation of autonomy. An approach that I regard as more useful – and one that I will take – will be to reframe the concern for moral autonomy posed by moral bioenhancement as an issue specifically of personal autonomy, and, to utilise insights from contemporary theories of autonomy to assess it on a deeper level.

5.3 Hierarchical accounts of autonomy

Contemporary accounts of autonomy have tended to associate the concept with the notion of authenticity. As mentioned above, this interpretation may be distinguished from the previously dominant Kantian conception of autonomy which posits that individuals are autonomous to the extent that their actions are motivated by a rationally driven will that is free from all contingent influences such as personal feelings, emotions, desires and inclinations. In contradistinction to this interpretation, contemporary interpretations of autonomy are more individualistic or phenomenological in their approach, as they acknowledge the importance of the first-person experience of autonomy. In other words, such a conception would regard individuals as autonomous to the extent that their "desires, actions, or character…originate in some way from [within, or from their] motivational set" (Stacey-Taylor, 2005:1). This conception of autonomy may also be conceived of as a type of "psychological property, the possession of which enables agents to reflect critically on their natures, preferences and ends, to locate their most authentic commitments, and to live consistently in accordance with these in the face of various forms of internal and external interference" (Piper, undated). In addition, most contemporary theories of autonomy are also deeply influenced by hierarchical accounts of autonomy, such as the independent theories of Harry Frankfurt (1971), Gerald Dworkin (1970), and Wright Neely (1974). Frankfurt's account of autonomy, which was briefly discussed in section 4.8.2, is informed by his interpretation of what confers personhood, where the latter is associated with those uniquely human qualities that human beings regard as not only the most salient aspects of who they are, but that may also pose potential challenges to their self-conception (1971:6).

Frankfurt's hierarchical account of autonomy[64] – first outlined in the 1970s – has had the strongest influence on subsequent theories of autonomy. The ability to form mental states such as desires and beliefs, and to act upon them, is, of course, not unique to human beings, as there is clearly the presence of such capabilities in other sentient creatures. Animals may not be able to form complex mental states such as beliefs, but they clearly exhibit states of desire. However, what is seemingly uniquely human, is the ability to assess and form opinions *about* our desires, beliefs and preferences, and, to therefore either endorse or reject them. In other words, we are able to distance ourselves from our seemingly lower or fundamental desires, which Frankfurt describes as first-order desires, and have higher or second-order desires about them, in a way that will either lead to us accepting and thus acting upon them, or will result in us rejecting them and wishing that we did not have such desires, even in cases in which we still act upon them. To use different terminology: I am able to want something (a first-order desire), and, I am able to *want* to want it (a second-order desire), or, I am able to *not want* to want it.

First-order desires encompass inclinations or intentions that one may possess but not be compelled to act upon. Such first-order desires are "mere impulses – pulls or temptations towards performing some action or other…[they may] be voluntarily adopted, but normally they simply arise unbidden in response to stimuli" (Ekstrom, 2005b:48). However, a subset of first-order desires is those desires that have sufficient motivational force to compel action. Frankfurt also uses the term *will* to refer to such first-order desires that are "effective…[in that they] move (or will or would move) a person all the way to action" (1971:8). In other words, a first-order desire that is strong enough to motivate one to act would be regarded as one's will[65]. Frankfurt uses the term second-order *volitions*, to refer to those second-order desires that individuals strongly desire to be their will, and thus, to motivate them to act. (Frankfurt, 1971:10). According to Frankfurt, true personhood requires that an individual must possess second-order volitions and not simply second-order desires. In other words, an individual must have the desire that one or another of his desires be his will, otherwise Frankfurt would regard him not as a 'person' but as a 'wanton' (1971:12). This will be explained further below.

Examples that illustrate the above distinction are abundant and several such examples were briefly discussed in section 4.8.2 of chapter 4b. However, now that a detailed investigation of Frankfurt's

[64] Frankfurt does not use the term *autonomy* himself, but rather refers to the *freedom of the will* (1971). However, his theory is taken as the foundation of most contemporary accounts of autonomy, and 'will', as he describes it, is synonymous with contemporary understandings of autonomy.
[65] From hereon I will therefore use the term *will* to refer to first-order desires that have sufficient motivational force to compel an individual to act upon them.

theory is being conducted, some additional examples may be discussed to illustrate the nuanced nature of his approach, and ultimately the relevance of his approach for the theory of autonomy that I will use to assess moral bioenhancement. An individual may have a will to regularly consume fattening foods, but at the same time, she may wish that her will was not effective in moving her to act, as she would like to be thinner and healthier. In other words, she may, despite her love of fattening foods, yearn to have a different will. She may wish to have a will that motivates her to be enticed by, and to consume, healthy rather than fattening food. In this regard, she rejects her will to consume fattening food, as it is at odds with her second-order volition that her will be to eat healthy food so that she can be slim. Another way of expressing this would be to say that while she has a strong desire to eat fattening food, she doesn't *want* to want to eat such food. She doesn't want her desire for fattening food to be the dominant will that motivates her to act, she wants her will to be different. Thus, when she acts upon her will to consume fattening food, she experiences distress, and in this regard her autonomy is compromised in some way.

We can also conceive of examples that have moral relevance. An individual may experience a strong emotional reaction of jealousy to his partner's interactions with others. This may elicit an overwhelming and effective will to control and restrict her interactions, in order to assuage his feelings of insecurity. When acting upon this will, it could then cause distress for both him and his partner, thus bringing conflict into the relationship. However, despite his strong will to keep his partner away from perceived threats to their relationship, he may at the same time wish that his will was different, that he was not this way and that he did not have such feelings. As was the case with the previous example, the individual rejects his will to control his relationship, as it conflicts with his second-order volition which is that he would rather act upon a will that is associated with being a trusting and caring partner. In other words, he doesn't *want* to want to control his partner, even though he *wants* to control her and finds himself constantly trying to. His will is at odds with his second-order volition; therefore, when he then acts in accordance with his will, he will experience distress, and once again, we could argue that his autonomy is compromised in some way.

A slightly different example would be an individual who recognises that she has a will to engage in extra-marital affairs. She is quick to experience boredom and enjoys the excitement of such illicit engagements; however, she wishes to keep the convenience of her marriage and enjoy its ensuing benefits. When critically assessing this desire, the individual may come to realise that any negative feelings she has towards her desires are not feelings of guilt, but are rather feelings of concern for any potentially negative consequences, such as being caught. In other words, she has

formed an opinion through consulting and reflecting upon her moral beliefs and has come to the conclusion that her will to engage in such affairs is morally acceptable to her. In this situation, her will is endorsed by a second-order volition to act upon her desires. She enjoys the desire that she experiences for her extra-marital endeavours and *wants* to want to experience such things. Thus, because there is congruence between her will and her second-order volition, regardless of the moral status of her desire, we could regard her illicit engagements as autonomous.

From the last example, it is therefore clear that, as is the case with most contemporary theories of autonomy, Frankfurt's account is *content-neutral*. In other words, on his account, an individual's autonomy is not predicated upon any particular conception of the good, but rather, upon what the individual regards as good. Content-neutral theories of autonomy may also be described as value-neutral. They specify procedures, processes or even particular structural aspects of an attitudinal state or action that must be present for an individual to be considered autonomous; however, they do not specify the nature of the actions or goals that individuals must enact or pursue in order to be considered autonomous[66]. Thus, an individual may have a will that motivates him to act in a way that is seemingly self-defeating, immoral or restrictive of his freedom; however, so long as he accepts or endorses his will with a second-order volition that is a product of critical reflection, then Frankfurt would regard him as autonomous. Describing second-order volitions as higher-order volitions, thus simply means that they are more closely aligned with the core self of the individual, or that the individual identifies more emphatically with such volitions[67].

Thus, we can revisit the example, discussed in section 4.8.2 of chapter 4b, of the unwilling drug addict. In such a case, the unwilling drug addict has a strong will to use drugs, but at the same time, she has an equally compelling and conflicting desire to not use them due to their deleterious effects on her life. However, she may find that her will possesses sufficient motivational force to compel her to use drugs, despite the fact that she has a second-order volition to abstain from drug use. In such a situation, because her will to use drugs is at odds with her second-order volition to

[66] Content-neutral theories may be contrasted with substantive theories of autonomy, which specify certain requirements for a decision or an action to be considered autonomous, where these requirements are informed by an underlying conception of the good. Kant's notion of autonomy would be an example of a substantive account. This is because it is informed by the underlying value system that he subscribes to. To be considered autonomous on a Kantian account, one has to act, or refrain from acting, in a particular way, that would only be regarded as autonomous if you agree with his conception of what is true and good. Content-neutral accounts, on the other hand, are attractive, as they have more universal applicability than substantive accounts. In other words, they are appropriate for use in societies characterised by value-pluralism and they are also, therefore, flexible enough to be used for contexts which require applied ethics approaches.

[67] In his hierarchical account, Dworkin describes this as being in a position to do one's "own thing" (1976:276), implying that one acts in an 'authentic' manner when one can form motivations to act, in a way that is informed and approved by one's own mental states, preferences and desires.

not use drugs, she would be regarded as having compromised levels of autonomy. We can also contrast this example with that of an individual who has a will to use drugs and simply acts on this will without reflecting upon the nature of his desire, such as whether or not he wishes his will to be effective in compelling him to use drugs or refrain from using them. As mentioned above, Frankfurt would describe such an individual as a 'wanton' (1971:12). Such an individual may even have conflicting first-order desires or inclinations, he may both desire and not desire to use drugs. However, in the absence of a second-order volition that one or the other of his desires be effective in compelling action, he would not possess personhood in the way that Frankfurt defines it. He would also be considered lacking in autonomy on Frankfurt's account. The unwilling addict, on the other hand, would, be considered to possess personhood on Frankfurt's account, due to the fact that he possesses a second-order volition to not use drugs. However, despite this, he would be accurate in considering himself to be lacking in autonomy in some way, as he will feel that he is being compelled to act against what is his will[68]

Thus, for Frankfurt, personhood is strongly associated with possessing the capability and the desire to form second-order volitions. Frankfurt argues that:

> it is only because a person has volitions of the second order that he is capable both of enjoying, and of lacking, freedom of the will…[Furthermore, personhood is not only constituted by possessing] both first-order desires and volitions of the second order. It can also be construed as the concept of a type of entity for whom the freedom of its will may be a problem (1971:14).

In the case of the wanton described above, without the presence of a will to *want* to want to use drugs or to *not want* to want to use drugs, the individual does not problematise the issue of whether or not his will is free, and thus, he is akin to an animal. Furthermore, as Frankfurt argues, an animal in the wild may roam freely; however, we would not regard it as having a will that is free in the sense that he describes. Thus, for Frankfurt, freedom of the will is different from freedom to act in one way or another, as he correctly points out that the latter can be restricted while the former remains intact[69]. Freedom of the will is being "free to will what [one] wants to will, or to have the will [one] wants" (1971:15). Thus, Frankfurt separates actions from the will that leads to them. Frankfurt argues that when there is congruence between second-order volitions and an individual's

[68] Frankfurt also considers those who have conflicting second-order desires to be on shaky ground – i.e. wanting to want something and not wanting to want something – as, in such a situation of conflict there will be no definitive second-order volition regarding precisely which first-order desire is favoured by an individual as her will. In such a situation, an individual will feel immobilised and unable to act (1971:16).

[69] This distinction between freedom of the will and freedom of action is, of course, the basis of Frankfurt's attempted refutation of the principle of alternate possibilities as a precondition for free will, discussed in section 4.8.4 of chapter 4b. In the example of the individual in the locked house, if she were to form a second-order volition that was congruent with her will to remain in the house – of which she is unaware is actually locked – then she would possess freedom of the will despite not possessing freedom of action.

will then her will is free, she is autonomous; when there is conflict between the two, then her will or autonomy is compromised to a degree[70].

These examples also illustrate the way in which hierarchical accounts of autonomy are strongly associated with authenticity. This is due to the stipulation that to be autonomous, an individual's will must be congruent with higher-level, second-order volitions, as the endorsement or rejection that is produced by the latter originates from the *self*. In other words, while we may feel that at times we are the captives of our first-order desires, due to their having the character of impulses, drives or instincts, because we are able to reflect upon and choose our second-order volitions, they seem to be more a part of our selves. Thus, second-order volitions give us hope that we have more control over our lives and our impulses than our first-order desires would imply; because of this, they feel more authentically ours than first-order desires. However, it must be noted that in terms of this process of reflection and selection, this would refer more to situations that are characterised by some form of overt conflict between different orders of desires, or by some aspect of the situation which draws attention to the need to reflect on second-order volitions. Frankfurt argues that, in general, congruence, or lack of congruence, between an individual's will and her second-order volitions occurs mostly in an unconscious manner (1971:17).

Frankfurt's theory, and hierarchical theories in general, are powerful for many reasons, one of which is the fact that they are easily able to justify the value that autonomy holds for us. In answer to the question of why we should value autonomy in an absolute sense, such a theory would be able to reply that autonomy is valuable due to the fact that it enables us to realise those desires that truly matter to us and are connected in a deeper way with who we take ourselves to be. On the other hand, if one's autonomy is thwarted, then one is impeded from realising these desires, and thus; from self-realisation. Furthermore, for Frankfurt, possessing both freedom of action and freedom of will represent complete freedom. He posits that where both are present, an individual "is not only free to do what he wants to do; he is also free to want what he wants to want" (1971:17).

[70] It must be noted that while Frankfurt has relatively stringent criteria for the presence of autonomy or freedom of the will, he most certainly does not argue that moral responsibility requires autonomy or freedom of the will. One of the purposes of autonomy theories is, of course, to serve as a means of articulating moral responsibility. In his attempted refutation of the principle of alternate possibilities as a precondition for free will, Frankfurt aims to contest the necessary link between autonomy and moral responsibility, thus arguing that an individual may be morally responsible for her actions even in cases in which her freedom of will was compromised or non-existent. I discussed this briefly in section 4.8.4 and will not address this area as it is unrelated to my area of focus in this dissertation. However, it is necessary to point out that Frankfurt has provided an extensive account of the subject of moral responsibility. See Frankfurt (1969).

Thus, by distinguishing between freedom of action and freedom of will, Frankfurt's theory gives credence to the intuition that what is important regarding autonomy is not simply captured by being free to act or do as one desires. This is because one's desires, and thus, the actions that they lead to, may be the product of manipulation or undue external influences, or they may be desires that we would rather not have. Furthermore, his theory – and hierarchical accounts in general – takes cognisance of the fact that autonomy is constituted not only by freedom from external coercion, but also by freedom from inner compulsions such as "phobias and addictions" (Ekstrom, 1993:600). In fact, inner compulsions and psychological disorders may constrain autonomy as acutely as external impediments[71]. While it will be shown below that Frankfurt's theory is vulnerable to the problem of manipulation, it is nevertheless a more comprehensive account of autonomy than preceding accounts.

5.4 Three problems with hierarchical accounts of autonomy

While hierarchical accounts of autonomy have laid the foundations for most subsequent autonomy theories, they are nevertheless subject to three problems. For an autonomy theory to be considered as a serious contender in the arena of applied ethics – and, in particular, to be useful as a means of addressing complex ethical situations that may arise in a biomedical context – it must be able to address and overcome these problems. Frankfurt has adapted his original theory slightly, in recognition of these problems, and subsequent theories of autonomy have employed different solutions to the problems, with varying levels of success. I will discuss the three problems, in order to illustrate how Ekstrom's approach successfully overcomes them, and thus, may be considered as a reliable and effective means of settling the concern for moral autonomy posed by moral bioenhancement. The first two problems are closely related but will be discussed separately as they are formulated somewhat differently and bring different matters of importance to the fore. The third problem has the most relevance for the concern for autonomy posed by moral bioenhancement.

5.4.1 The problem of infinite regress

On Frankfurt's hierarchical account, autonomy is secured when an individual's will is endorsed by a second-order volition. As mentioned above, the latter is regarded as having more authority than

[71] The way that one might be constrained by various addictions such as to a particular substance, to a toxic relationship, or by compulsive behaviour, has been described as "the internal problem" by Lehrer due to the fact that such internal constraints are as inhibiting as external constraints (Lehrer, 1997 in Ekstrom, 1999:1058). Hierarchical theories of autonomy are, of course, a recognition and response to the specific nature of this problem.

the former, and is thus regarded as more closely associated with the self. However, one could, of course, question the validity of second-order volitions. In particular, one may ask why an action that has been endorsed by a second-order volition is more autonomous or authentic than one that hasn't been. In other words, one can question the source of the supposed reliability or authority of volitions of the second-order and ask from whence this originates. In order to answer this question, one would have to move backwards and possibly affirm the reliability of one's second-order volition by pointing out that one has a positive disposition towards it. This positive disposition then becomes akin to a third-order volition. However, the problem persists; requiring one to move further backwards in an infinite regress of justification.

To illustrate with an example, I may be a person who has no desire to be generous and assist those in need. My will is such that I never donate to charities or provide any form of charitable assistance. When assessing my will, I experience feelings of guilt and wish that I had a strong desire to be generous, that would motivate me to act accordingly. In other words, I wish that I wanted to be generous, or, I want to want to be generous. This second-order volition would be regarded as indicative of what I truly want, as I have clearly thought about the matter, and it would thus possess authority to 'speak' for me. However, I could also assess this second-order volition by means of a third-order volition to be more accepting of my shortcomings and not so ready to feel guilt for my lack of generosity. In other words, I could wish that I didn't want to want to give and were not plagued by guilt. I could then assess this third-order volition by wishing that I didn't wish to be more accepting of my shortcomings: I don't want to not want to want to want to give. This regress could then continue backwards ad infinitum. In this example, there is conflict between different levels of desire, however, we could also come up with more straightforward examples. I could have a desire to do something and this could be endorsed by wanting to want it, which could be further endorsed by wanting to want to want it and then by wanting to want to want to want it, and so forth.

Frankfurt does recognise that there is a potential problem of regress, particularly in the case of conflicting levels of desires, as described above. He admits that "[t]here is no theoretical limit to the length of the series of desires of higher and higher orders; nothing except common sense and, perhaps, a saving fatigue prevents an individual from obsessively refusing to identify himself with any of his desires until he forms a desire of the next higher order" (Frankfurt, 1971:16). In other words, while Frankfurt recognises that we could relentlessly reflect on our desires, he believes that doing this would be a decidedly negative thing. Furthermore, he posits that the solution to the

problem of infinite regress would be to halt this process in a non-arbitrary manner by *decisively* or *resoundingly* committing to one's will. This may be possible in cases characterised by minimal conflict between levels of desire, such as a single-minded desire regarding an endeavour for which there is no presence of self-doubt. However, in a great deal of cases, such as the example of the ungenerous individual discussed above, it is not self-evident that halting the process of reflection at a particular point would escape a charge of arbitrariness. Critics therefore regard Frankfurt as having failed to adequately address the problem of infinite regress[72].

5.4.2 The problem of authority

The second problem has been referred to by different terms but it encapsulates the issue of how second-order volitions are deemed to be higher or more authoritative. In this regard, it is closely related to the problem of infinite regress. In other words, while occurring on seemingly different levels, desires are nevertheless the same phenomena, they are all desires (Watson, 1975:218). How then, may we argue that one desire should assume primacy over another? Another way of formulating this problem is to frame it as one of identification. In other words, one may ask why an individual *identifies* more strongly with a second-order volition than with a first-order desire[73]. As discussed above, the solution to this problem would generally be to posit that second-order volitions are more authentic in their closer association with the self because they are more considered than first-order desires; but how do we know that one desire is more one's own than another? One could reject a first-order desire that is at odds with a second-order volition because one is in a state of self-denial about what one really wants. In other words, one could be in denial regarding one's true second-order volition. Because Frankfurt stipulates no substantive conditions

[72] It is interesting to note that it is not only in the area of volitional justification that we find this problem of infinite regress. This is, strictly speaking, also an epistemological problem that characterises any knowledge claim, grounded in a classical model of rationality (Van Niekerk, 1980 & 1983). Albert (1968) describes this problem as the Münchhausen trilemma, referring to Baron Münchhausen who was required to drag himself out of a swamp by his very own hair in a classic example of bootstrapping. More specifically, the Münchhausen trilemma refers to the problem of how we ground the truth of our knowledge claims in accordance with the classical model of rationality. Cambier (2006) describes this as Albert's identification of "the problem that rationalist philosophy cannot itself establish its own foundations" (2006:145). Generally, the condition for a proposition to be regarded as a knowledge claim is its ability to be grounded through sufficient reason, or by way of some form of proof. However, one could then enquire as to the veracity of this grounding or proof, which according to this model, requires grounding or proof itself, and so forth. In response to this problem of seemingly infinite regress of justification, Albert argues that we are faced with a trilemma: we can accept and engage in this infinite regress of justification, we can employ a circular form of deduction whereby we use proof that itself requires justification, or we can simply halt the process in an arbitrary and dogmatic manner (Van Niekerk, 1983:14-29).

[73] The problem of authority is also sometimes referred to as the problem of identification for this reason. The issue at hand is: what does it mean to identify oneself with a desire that one has, so that in doing so, any act that ensues is self-determined.

for the way in which second-order volitions should arise or be formed, his theory is not able to adequately address this problem.

As mentioned above, Frankfurt's account is content-neutral, therefore, when an individual assesses her will she does not necessarily do so by taking a "moral stance" (1971:13) towards it. Frankfurt argues that:

> a person may be capricious and irresponsible in forming [her] second-order volitions and give no serious consideration to what is at stake. Second-order volitions express evaluations only in the sense that they are preferences. There is no essential restriction on the kind of basis, if any, upon which they are formed. (ibid.)

The problem, however, is not the fact that second-order volitions must be assessed in terms of some moral position, or that desires must be 'pure' in some way, rather than informed by internal volitions; rather, what is problematic, is that in his lack of restrictions regarding how desires may be formed to be taken as higher-order or authentic, Frankfurt's account is once again plagued by arbitrariness.

As discussed above, in his initial paper on freedom of the will, Frankfurt attempted to pre-empt this concern to a certain extent by positing that second-order volitions are higher, and thus, able to halt the regress backwards, due to the individual's decisive commitment to them (1971:16). However, as mentioned above, this response was not regarded as having successfully addressed the problem (Watson, 1975:219). In a later chapter, Frankfurt addresses the problem further, arguing that decisive commitment to a second-order volition does not represent an arbitrary drawing of the line. Here, he draws an analogy with the calculation of a mathematical problem. An individual makes a calculation and then checks her calculation for accuracy. She can continue to check her calculation again and again, but at some point, she will stop her checks for the reason that she is "unequivocally confident" (Frankfurt, 1988:168) that they are, in fact, accurate, and that no further checks are necessary. In this situation, the individual believes that the answer to the mathematical problem will remain the same with each future check, and thus, her "commitment resounds endlessly…[and is therefore] made without reservation" (ibid). In a similar manner, a person will check his second-order volitions, and, in cases where there is conflict will check until he has eliminated any self-doubt. Frankfurt's argument is that stopping the process due to having eradicated all self-doubt, and thus, trusting in the fact that one's higher-order volition is truly reflective of one's desires is not an arbitrary halt of the process. The person will have no valid reasons to continue the process. When reaching a position where there is an absence of conflict in the accuracy and authority of a

higher-order volition, the individual can then be said to be *wholehearted* in his commitment to it (1988:175).

However, despite Frankfurt's response to this problem, critics have remained unconvinced. While one can wholeheartedly commit to a desire and stop the process implying an authority of sorts, the problem of arbitrariness nevertheless remains. This is because without any criteria by means of which one can defend one's second-order volitions, in terms of indicating why they are higher, they simply become desires that one happens to have (Ekstrom, 1993:602). If one frames this problem in terms of being one of identification and argues that second-order volitions are of a higher-order because the individual identifies more strongly with them, and thus, that they are constitutive of her true self, then this implies that first-order desires that are at odds with second-order volitions are distinct from, or not a part of, the true self. However, because Frankfurt has provided no account of the self, in terms of constructing a theory regarding two important requirements, namely, identifying "[w]hat is essential to a self…[and secondly w]hat is it for a self to identify with some desire, course, of action, or belief, deeming it as one's own" (ibid. 603) he cannot evade the charge of arbitrariness[74].

In terms of determining the essential nature of the self that identifies with a particular second-order volition, and thus gives it its authority, this is not a subjective matter that is left up to the individual in question. In other words, the individual does not simply decide what characteristics are integral to selfhood; rather, this must be supported by a theory of selfhood. It is only in the case of the second requirement, the act of identifying or not identifying with a desire as *one's own* that the matter is left up to the individual. To explain the difference between the two, Piper refers to the distinction between "self-conception and conception of the self" (1985:174), where the former refers to subjective components and the later to the objective components. On this distinction, if I have a belief that I am a morally upstanding individual, this would form part of my self-conception, whereas the question of whether or not an ability to act in accordance with moral beliefs is essential to selfhood, would form part of the conception of the self, and thus, would not be a matter that is left up to me. Therefore, we require a more substantial or definitive way of clarifying whether my seemingly higher, second-order volitions that motivate me to act are truly my own, and are thus autonomous, in terms of the latter implying that they are *self-determined* (Ekstrom, 1993:603).

[74] Because the concept of autonomy is generally taken, at its most fundamental level, to signify the notion of self-determination, this means that a coherent autonomy theory – particularly one that wishes to avoid the problems discussed here – must provide some theory on the self so that it can clearly stipulate what it means to act in accordance, and thus freely, with the self. This is a point that Dworkin has also raised (1976:23).

5.4.3 The problem of manipulation

Even if hierarchical accounts of autonomy, such as Frankfurt's, are able to address and successfully overcome the two afore-mentioned problems, there is a third and more serious problem that remains, which also happens to be of crucial importance for the matter of the concern for autonomy posed by moral bioenhancement. Of course, if autonomy is predicated upon the ability of the self to determine its own course, subject to specific criteria depending on the nature of the theory, then a prerequisite would be ensuring the absence of undue external constraints or influences which could thwart this ability. External constraints include not only obvious instances of constraint such as overt coercion or force utilised to compel an individual to act in a particular manner, but also include more subtle forms of manipulation and coercion that could impact upon the formation of desires, preferences, beliefs and other mental states. This would include impacts on the above mental states produced by neurobiological interventions. To be regarded as a serious contender, therefore, an autonomy theory cannot simply provide content regarding the relationship between the will, or motivation, and second order volitions. It must be equipped to deal with cases in which this relationship may happen to be in harmony purely as a result of undue external influences.

Hierarchical accounts of autonomy, such as Frankfurt's, are particularly vulnerable to the problem of manipulation, due to the fact that they are ahistorical or purely structural accounts of autonomy. In the case of Frankfurt's account, he examines the relationship between the will and higher-order volitions and argues that this relationship must possess a particular structure for the individual in question to be considered autonomous. However, there are no stipulated requirements outside of this structure, in terms of how an individual has come to acquire his desires and second-order volitions, that preclude them from having been imposed by external sources. In other words, in situations in which an individual's higher, second-order volitions have been manipulated by a hypnotist, through behavioural modification techniques, subliminal suggestion or through a neurobiological intervention, Frankfurt's account of autonomy would not be able to elucidate why this would be problematic for autonomy as he gives no suggestions regarding the process by which second-order volitions must be formed[75].

[75] For this reason, certain theorists have developed explicitly historical approaches to autonomy, a noted example being the theory of John Christman (1989) who argues that an individual must accept the historical process by which she came to have particular desires and preferences (in Stacey Taylor, 2005:2).

Dworkin's earlier hierarchical account gives more substance to the notion of autonomy than Frankfurt's but is also vulnerable to the problem of manipulation. His formula for autonomy is: "autonomy = authenticity + independence" (Dworkin, 1976:24). Furthermore, he is also of the view that autonomy is present on the second-order level rather than on the first-order level. There is very little autonomy at the first-order level due to the fact that as individuals, we are shaped and influenced to a major extent by socio-economic, cultural, environmental, psychological and other biological factors, over which we have minimal control. Thus, Dworkin argues that our will, or our "beliefs, desires, emotions, principles, and so forth" (ibid.), cannot be accurately described as having been freely chosen or adopted. However, as he points out, while we have little autonomy in choosing the above-mentioned mental states, we do have autonomy in our ability to assess and adjudicate them.

Dworkin has refined his views, in subsequent publications, to ensure that they are even less susceptible to the problem of manipulation. He had originally used the notion of *identification* of second-order volitions with first-order desires as a criterion for authenticity and thus autonomy. However, he abandoned the term identification in favour of the requirement that autonomy is secured by having the *capability* "to raise the question of whether I will identify with or reject the reasons for which I now act" (1988:15). His change of heart in this regard was motivated by his belief that autonomy is a concept that must be measured over a substantial period of time, rather than localized at a particular point in time. In other words, desires, and identification with them, must be looked at over time and autonomy lies in an enduring congruence between the two. Furthermore, because an individual's identification with her first-order desires may be the product of manipulation, what is more relevant, is thus whether or not the individual possesses the *capability* to identify with, or, to renounce a first-order desire. If she has been manipulated in some way she would not truly possess such a capability.

Reframing the problem in this way gives a definitive solution to certain counter-intuitive examples. One such example is that of a willing drug addict; an individual who is accepting of his addiction and desires it to be no different from what it is. On Frankfurt and Dworkin's earlier account, because the willing drug addict has a second-order volition that identifies with his first-order desire to use drugs, he would be regarded as autonomous. Dworkin's reformulation enables him to identify why the willing addict should not, in fact, be regarded as autonomous. This is because the addict may have deliberated on his desire to use drugs and come to identify with it, but in reality, his addiction is such that he lacks the genuine capability to change his desire. Thus, Dworkin comes

to the conclusion that the notion of autonomy does not simply require that one adjudicate or judge one's preferences, but should also require that one be able to change one's preferences and act upon them (1988:17)

Dworkin also adds a further condition for autonomy; he stipulates that the process by which an individual comes to acquire his desires or preferences must have procedural independence. That is, an individual cannot have come to acquire his desires or preferences in such a way, or through such a process, that would be regarded as having been unduly imposed on him (Dworkin, 1976:25). Dworkin distinguishes between procedural independence and substantive independence. Procedural independence is characteristic of a content or value neutral approach to autonomy, as discussed earlier, as this account looks at the process or structure of the forming of desires and other motivational states in order to determine whether they are the individual's own. It also distinguishes between acceptable or unavoidable influences, and problematic or avoidable influences on our independence. Substantive independence, on the other hand, is characteristic of a substantive approach to autonomy and would thus place restrictions on which acts would be deemed to be autonomous. To maintain substantive independence, an individual must have been acting in an autonomous way when her motivations were formed. In other words, as he argues, an individual cannot have "renounce[d her] independence of thought and action" (Dworkin, 1976:25) before she formed her motivations. She may authentically give away her autonomy on a procedural account; however, on a substantive account this would be paradoxical[76].

The problem of manipulation is, therefore, an issue of authenticity. In this regard, some autonomy theorists have stipulated that to be autonomous, two basic requirements must be fulfilled: actions are subject to "*authenticity* conditions and *competency* conditions" (Christman and Anderson 2005:3). Authenticity conditions, refer to a conception of autonomy that is espoused in hierarchical theories such as those of Frankfurt (1971) and Dworkin (1970), where authenticity is predicated upon the extent to which one's desires, preferences, beliefs, etc., are a product of one's own deliberation and are accepted as one's own. Competency conditions, on the other hand, are specific capabilities that must be present for an individual to be autonomous. These include the capability for "rational thought, self-control, self-understanding, and so on" (Christman & Anderson,

[76] This account is fraught with difficulties however, as it would argue that one gives up one's substantive independence by wanting to do what one is told to do, or by being overly concerned with the dictates of peer pressure, for example. However, if this is accurate, then we are forced to admit that "the compassionate or loyal or moral man is one whose actions are to some extent determined by the needs and predicaments of others" (Dworkin, 1976:26), and thus, that he lacks self-determination. Elsewhere, Dworkin posits that this view of autonomy "seems in conflict with emotional ties to others, with commitments to causes, with authority, tradition, expertise, leadership and so forth" (1988:7). This is because one's commitment to these factors impact upon one's ability to self-determine.

2005:3)[77]. Furthermore, and relevant for the problem of manipulation, individuals must not only possess these capabilities, they must also be able and "free to exercise…[them], without internal or external coercion" (ibid.). However, while authenticity and competency conditions may be necessary for autonomy, they are not sufficient, as an individual could meet both conditions but not be truly autonomous, for a number of reasons. An individual may have formed her desires and preferences in an authentic manner, without possessing the relevant facts due to having been manipulated in some way, or, she may have been hypnotised or 'brainwashed'. Or, most importantly for the purposes of moral bioenhancement, she may have undergone some neurobiological intervention that alters the means by which she identifies her desires and preferences, so that she now identifies with desires and preferences that she had previously rejected before the intervention.

This is the same concern that Dworkin has, which was mentioned on the previous page. He argues that autonomy is not realised simply due to being able to form an opinion regarding one's preferences. Rather, a precondition for autonomy should be the possibility that one be able to change one's preferences and act upon them (1988:17). This is also the specific concern that is voiced by Harris in response to the claims of proponents of moral bioenhancement who insist that rational or deliberative capacities will not be circumvented by moral bioenhancement interventions as Harris fears. What Harris is specifically concerned about, is not the ability of these capacities to function, as it seems obvious that receiving moral bioenhancement would not render an individual incapable of deliberation and the ability to employ rational capacities; rather, it is the possibility that moral bioenhancement could alter the specific conclusions that these capabilities would be likely to enable individuals to come to, due to altering their desires and preferences. This is essentially, the nature of the problem of manipulation.

The relevance, then, of procedural independence, and thus, of the problem of manipulation, for moral bioenhancement, is to identify which external forces and influences alter an individual's judgements and ability to assess her desires, preferences, beliefs, choices and actions to the extent that we would no longer describe her assessments and judgements as being her own. This is not an easy task due to the fact that, as alluded to by Dworkin, in assessing when procedural independence has been impacted upon we must distinguish between acceptable and unacceptable influences on "higher order judgements" (1976:26). In other words, we must be able to explain why some

[77] Feinberg has also discussed the notion of competency conditions which he refers to as "autonomy as capacity" (1989:28).

influences, such as socialisation, peers, upbringing, education, cultural background etc. are regarded as acceptable influences on our ability to assess our mental states, while other influences, such as moral bioenhancement, would be problematic.

5.5 Criteria for an adequate theory of autonomy

In this section, I will present six criteria that have been devised by Dworkin to assess the efficacy of a potential theory of autonomy (1988). Dworkin does not posit that all these criteria could necessarily be met by a theory; however, a theory that could meet as many of these criteria as possible would warrant serious attention in terms of its ability to be utilised to navigate difficult cases in the arena of applied ethics, for example. In the following section, I will present Ekstrom's coherence theory of autonomy, which I argue is up to the task of navigating the concern for autonomy posed by moral bioenhancement. After this, I will provide a justification for this argument by illustrating how Ekstrom's theory meets Dworkin's criteria as well as how it overcomes the three above-mentioned problems.

Dworkin argues that to be considered as an adequate account of autonomy, which may then be utilised as an effective tool of assessment, a theory must meet certain requirements. The first requirement is that a theory of autonomy cannot contain any logical inconsistencies in its particular interpretation of autonomy or its use of related concepts (Dworkin, 1988:7). The example given to elucidate this requirement is the notion of "an uncaused cause" (Dworkin, 1988:7). If a theory were to stipulate that autonomy is predicated on this kind of radical indeterminism, where such a phenomenon is regarded as logically incoherent, then the theory in question would fail to meet the requirement of logical consistency.

Secondly, an account of autonomy cannot interpret the notion in such a way that autonomy becomes impossible to possess in a practical sense (ibid.). In other words, if a theory has overly strict requirements that result in it being impossible for anyone to be, or ever have been, autonomous, it would fail to meet this requirement. An example of such a theory would be one that stipulates that autonomy requires that one's moral code be entirely the product of one's own volitions, and not have been influenced in any way by external sources such as upbringing and socialisation, socio-economic position, and so forth. The influence of such sources is generally regarded as inevitable, and thus, a theory that seeks to exclude such impacts would be impractical. On this interpretation, autonomy would be impossible to achieve.

Thirdly, a prospective theory must be able to indicate why autonomy is regarded as a good thing, or as something that is worth having (Dworkin, 1988:8). A theory could do this in various degrees of strength. It could argue that autonomy is of absolute or ultimate worth, more valuable than any other good, or it could make a weaker claim that it is one, amongst many, goods. To this requirement, Dworkin adds that a theory should not imply that autonomy may only be achieved at the expense of other 'goods' regarded as valuable, such as justice, loyalty or equality. This latter point is important because substantive accounts of autonomy have been criticised on the basis that they are incompatible with certain values, as mentioned in footnote 76. While Dworkin does not allude to this, elsewhere he has criticised and rejected substantive accounts for various reasons, arguing rather for the desirability of content-free accounts of autonomy (1988:21-25). Presumably this criterion is included to address this issue.

The fourth requirement is that a theory of autonomy should be one that can be utilised regardless of the value system in question. In other words, it should have *ideological neutrality* (Dworkin, 1988:8). While Dworkin stipulates that this is not a strong requirement, what he means by this is that a prospective theory should frame autonomy in such a way that it has broad appeal across a variety of world-views and belief systems. As he points out, a conception of autonomy should be applicable to a diversity of outlooks ranging from those that espouse strong individualism to those that balance autonomy with other values.

The fifth requirement is that a prospective theory of autonomy must have practical and normative applicability. In other words, it must be possible to use it in a philosophical context as a means of arguing for or against phenomena that are associated with different freedoms or regarded as inimical to them. An example here, would be using a particular theory of autonomy to critique an overly strong interpretation of autonomy that rejects any form of state intervention as a violation of autonomy, thus supporting anarchy. Another example would be using a theory of autonomy to illustrate why paternalism in medicine is problematic.

The final requirement is that of *judgemental relevance* (Dworkin, 1988:9). What is meant by this stipulation is that an account of autonomy must be congruent with generally accepted claims regarding autonomy. One such claim would be to interpret autonomy either as a threshold concept or one that admits of degrees. Another such claim would be to associate autonomy with an opposition to paternalism, while yet another would be the view that to inculcate the value of autonomy, one must introduce children to the notion in increasing incremental stages. These claims

are examples of "judgements that are conceptual…normative…[and] empirical" (Dworkin, 1988:9) respectively. I will return to these criteria in section 5.7, once I have presented Ekstrom's theory and utilise them to assess her theory.

5.6 Ekstrom's coherence theory of autonomy[78]

5.6.1 Background information

Ekstrom's coherence theory of autonomy is also based upon the canonical interpretation of autonomy as implying self-determination, where this latter interpretation means acting in accordance with one's "own reasons" (1993:599). However, as she explains, to build a comprehensive theory out of these insights, an account must firstly provide conditions for *how* the reasons that one acts upon, in the case of autonomous acts, are one's own, or not, in the case of non-autonomous acts. Secondly an account must explain what is meant by the notion of *one's own*, by providing a theory or account of *the self* that can explain what is internal and external to the self (Ekstrom, 2005b:52).

Ekstrom's approach is strongly influenced by hierarchical accounts of autonomy as she builds her theory upon the view that individuals not only have desires, attitudes, beliefs and other mental states, they also have the ability to have opinions and preferences *about* the afore-mentioned. The enduring significance of the hierarchical approach is due to its many strengths and intuitive appeal, one of which is that it clearly explains why, and how, conflicts between these two levels impact upon the autonomy of the individual. Thus, it indicates why autonomy is not only vulnerable due to external sources or threats, but also due to internal conflicts (Ekstrom, 2005a:143). These internal conflicts can produce such distress for individuals, that they result in feelings of self-alienation (2005a:146). In certain cases, an individual may feel disconnected from his own desires or actions and experience feelings of repugnance towards them. While different hierarchical approaches would explain the nature of these conflicts in different ways, their value lies in having actually identified this additional internal source of conflict as an impact on autonomy, one that may be equal in force to external impacts on autonomy.

Furthermore, a theory such as Frankfurt's is attractive due to the fact that the identification of a second-order level of judgement, as a means of assessing our impulses and drives, gives us the

[78] The explication in this section is based upon an amalgamation of Ekstrom's position as outlined in several of her publications (1993, 2005a, 2005b).

sense that we are not purely at the mercy of the latter. Through some ability that is internal to the self, and thus perceived as authentic, we may exert control over our baser impulses and drives. As Ekstrom posits, this is a valuable capacity because "[i]f one is able to get oneself to want what one wants to want, and one is able to get oneself to act as one wants to want to act, then it seems that one has achieved some control over both one's mental life and one's actions in the outer realm" (2005b:49).

However, what Ekstrom takes to be important for autonomy, is not the view that autonomous action is *solely* predicated upon acting in accordance with our second-order volitions, or acting "from a desire for another desire" (2005a:147). Rather, for her, what matters for autonomy is *how* higher, second-order volitions are formed, namely, through the process of *reflective endorsement* (2005a:147)[79]. Through this ability, we can critically analyse our mental states and actions in terms of how valuable they are for us to have and to act upon, and this assessment is not only performed by the self but is also instrumental in forming the self (ibid. 148). However, simply framing the issue of autonomy in terms of a distinction between high and low-order desires and volitions, as Frankfurt does, produces problems of the sort described in section 5.4 which require that more substance be added to such an account. Ekstrom's theory, which specifically avoids the notion of higher, second-order volitions, aims to avoid the problems that plague such accounts, and I would argue that it does so successfully.

5.6.2. Preferences

Ekstrom's account of autonomy diverges from the traditional hierarchical approach at the outset by utilising the notion of *preference* rather than desire, as a basis for her theory. She defines the theoretical approach of coherence theories of autonomy as requiring "that the preference on which a person acts must cohere with other attitudes, in order for the act to be autonomously performed" (Ekstrom, 2005a:144). A preference is a specific type of desire that bears similarities to a Frankfurtian second-order volition in that it is a desire for a desire on the first-level to be "effective in action" (1993:603). However, Ekstrom adds the qualification that a preference is directed towards achieving 'the good', for the individual in question, in some way. In other words, an individual forms a preference to act on one of the first-level desires that she has because she has assessed her first-level desire against a conception of what she takes to be "true and good"

[79] As Ekstrom points out, Frankfurt does not include the requirement of reflective endorsement as integral to the process of adjudicating our desires as he would regard such a criterion as "excessively rationalistic" (Frankfurt, 2002:89). He would simply stipulate that our desires must be accepted wholeheartedly to count as second-order volitions.

(2005b:55). This process may or may not be consciously performed, however, the important point is that, on her account, there is some form of process that is required to form a preference (2005a:148). By the term preference, Ekstrom does not refer to a comparison, whereby an individual prefers one thing to another; rather, she uses the term "stipulatively…[to refer to] a desire that has survived a process of critical evaluation…with respect to an individual's conception of the good" (2005a:148). A preference, so defined, is "an evaluated desire" (Ekstrom, 2005b:54).

As human beings, we are constantly forming and acting upon different desires for various reasons that are informed by the kind of person we are. We also find ourselves having momentary desires, akin to whims, that we are not be able to explain with reference to the kind of person we are. These may be short-lived desires that suit us to have at a specific time, or they may be desires for something, or some course of action, that arise due to internal pressures such guilt, or external pressures such as peer pressure. They may even be the product of instincts. For Ekstrom, however, preferences refer only to those desires that are informed by the kind of person that we are – our character – and not the fleeting kind of desires.

As mentioned above, while Ekstrom's notion of a preference is similar to Frankfurt's notion of a second-order volition, it is more rigorous due to her stipulation that preferences are formed due to an individual's consideration that a first-order desire is good or will bring about 'the good' in some way. A Frankfurtian desire is considered a desire simply because an individual has it; there are no requirements regarding how it may be formed, thus leaving his account open to the threat of external manipulation. Furthermore, Ekstrom doesn't describe preferences as being of a higher level themselves; rather, she posits that they are the products of "higher-order mental states" (1993:604), due to the fact that they are produced through a process of critical reflection "with respect to the standard of goodness" (605). There is strong empirical evidence for the existence of such higher-order mental states. An example would be an individual who has a wish to give to the poor but also wishes to do so because it is the right thing to do and wants it to give him pleasure to do so, rather than doing so reluctantly or simply to assuage feelings of guilt. In other words, he wants to give to the poor but he wishes to do so for reasons that are considered worthy or good. This latter ability to assess and to come to such conclusions is indicative of a higher-order mental capacity[80].

[80] It must be noted that this should not be taken to indicate that a higher-order mental capacity implies that the individual acts upon reasons that are more morally worthwhile in an objective sense. Worthwhile here, simply implies valuable for the individual in question. This matter will be explained further below.

5.6.3 A theory of the self

As alluded to above, to give an account of autonomy as self-determination that avoids both the problems of infinite regress and authority, an autonomy theory must be able to adequately explain what it means for preferences to be one's own or to come from one's self. This, in turn, requires a theory of the self. There are many ways in which the matter of defining 'the self' can be approached. As Ekstrom posits, one may approach the issue from a "metaphysical, semantic…[or] epistemological" (2005b:45) paradigm. She is careful to point out, however, that by providing such a theory, she is not attempting to give a "metaphysical account of personal identity" (1993:600) or a Cartesian notion of the self. She is engaging, rather, with a "a moral or psychological self" (Ekstrom, 2005a:149). The popular understanding of the notion of self is generally encapsulated by a reference to what is mine: *my* thoughts, *my* beliefs, *my* preferences, *my* actions, and *my* physical body. However, there are frequent occasions in which we feel alienated from our own thoughts, beliefs, preferences, actions and even our own bodies. In such instances, it makes sense to describe them as *not mine*. It is this psychological or moral account of the self that is relevant for the issue at hand.

While Frankfurt does not provide a specific theory of the self, the implicit understanding that informs his account is that the self simply *is* the higher, second-order volitions or desires of the individual. In this regard, he gives no importance to the role of rationality – or any other factor, for that matter – in the formation of desires, and thus, as part of the self. Ekstrom concurs with Frankfurt that autonomy is informed by the extent to which one can act upon desires and other attitudinal states that one has, rather than upon "judgements" (2005b:53); however, she regards the role of "evaluative reasoning" (2005b:53) as pivotal in this process of desire formation. This is because she argues that in deciding which desires will move an individual to act, the autonomous individual forms his choices through evaluating them in accordance with some standard that is internal to him.

Ekstrom's interpretation of the self is, therefore, a broader conception, where the self refers to a "particular character" (1993:606) as well as the uniquely human evaluative ability to form, mould and continually adapt that character. A character, or *character system*, refers to the "aggregate of preference and acceptance states" (2005b:54), or attitudinal states, that have come to be acquired by an individual through a process of evaluation and deliberation. In other words, by being part of her character, these preference and acceptance states have been measured by the individual against her standard of what she takes to be good and true, and endorsed as being acceptable in moving her

to action. Ekstrom does not regard the sum total of all desires and beliefs that individuals have as part of their character, however; her conception is narrower than this.

As we live our lives through time, our desires and beliefs are subject to various changes and vagaries, some of which we are willing to accept and others that we are not. In this sense, our character refers to our delimitation as a particular self that is recognisable through time to others and can be distinguished from others. While any fleeting and rejected desires and beliefs are mine in the sense that I am aware of them as I experience them, they are not an integral part of my character if I have not *accepted* them. Lack of acceptance in this regard would indicate that they do not measure up to a conception of the good that I have, and thus, that I would not wish them to be the basis by which I am moved to action. In other words, if I were to give an account of my character, I would not include such beliefs and desires in my account of who I take myself to be. The critical assessment by which this endorsement or rejection is conducted, is internal to me. Thus, while I cannot control the fleeting first-order desires that arise within me, and that I do not identify with, I am able to control which desires and beliefs become preference states that I would wish to move me to action. This is what Ekstrom means by a self, or, a "moral or psychological identity" (2005b:55). It is not only the preferences – understood in her stipulated interpretation as desires that have been critically evaluated against a conception of what is true and good – that have come to be accepted by the individual that constitute her self, character or identity, it is also her ability to actually form these preferences and thus, her ability to form her character.

One could respond to the above claim by arguing that it is an idealised account of character in that an individual's unwanted first-order desires and beliefs are as much a part of his integral character as his higher-order aspirations. Such first-order desires could, in fact, be more indicative of his character than the latter. Ekstrom's account of character could rather be a case of wishful thinking; it could be more indicative of the character one would prefer to consider oneself to have than the one that one actually has. However, Ekstrom foresees such a response and points out that the way we speak about ourselves in everyday life lends support to her claim. We often find ourselves having a belief or desire, or behaving in such a way "in spite of ourselves" (1993:607). In other words, while our characters inform what we take to be true and good, we sometimes have a desire that causes us to behave in a way that is at odds with our conception of the good[81]. Such an

[81] Ekstrom's conception of what is true and good is, of course, a personal conception of the good, thus, it is important to bear in mind that the nature of desires and beliefs that are considered anomalous could be positive or negative. In other words, an individual who has an objectionable character and is considered to have selfish preferences by others, could find himself having "an objectively noble impulse in spite of [himself]" (1993:608).

occurrence may cause us to comment that we are not ourselves, or it may result in others observing that we are behaving 'out of character'. If such behaviour is enduring enough, it may then be the basis upon which others notice that we have changed and are now not the person we once were. In the case of anomalous desires and beliefs, we may believe that these alternate states are part of us in some way, but on Ekstrom's account, if we don't identify with them we will not endorse them, and, most importantly, we wouldn't be prepared to defend them (1993:608). For a preference state to be a component of an individual's character, an individual must be able to explain or give reasons for this preference.

5.6.4 Coherence, personal authorisation and the integral-self

Thus, Ekstrom has described the self or the character, as informed by the ability of the individual to deliberate upon and endorse certain enduring preferences that accord with her conception of the good. However, to avoid giving an account of autonomy as self-determination that avoids the problems of regress and authority, what remains is for Ekstrom to indicate why the process whereby a desire becomes a preference has sufficient authority "to stand for the perspective of the agent – to indicate what 'she really wants' or to count as the desire with which she identifies" (2005b:58). In order to achieve this aim, she deepens her account further to include the notion of a "true or most central self" (1993:608) which is a component of an individual's character[82]. This integral self consists of the evaluative character forming ability that individuals possess, along with only those "acceptances and preferences…that cohere together" (ibid.), where coherence refers to the way in which these different attitudinal states work in conjunction with one another and are connected in a way that is stable, harmonious and able to be explained with reference to one another. In other words, cohering preference states form an interrelated "structural arrangement" (2005a:149). Ekstrom also introduces the additional requirement that preferences must be *authorized* by the individual, if they are going to be the basis upon which she acts in an autonomous manner, and argues that this occurs when the individual has critically examined a particular preference and found it to be coherent "with [her] other preferences and acceptances" (ibid.). Having a cohering network of preferences, therefore, serves to provide an individual with good reasons as to why he accepts or rejects a particular preference and thus chooses to act upon it or not.

Thus, in terms of the problem of infinite regress, we do not need to turn to a higher-order desire to endorse a preference we have; rather, we need to have critically assessed a preference's worth in

[82] I will refer to this notion of the true or most central self as the *integral self* from hereon.

terms of whether or not it is defensible by means of the set of cohering preferences that form our integral self. In other words, examining a preference to assess and justify it as one's own would require that one be able to defend it with reference to the fact that it is coherent with one's integral self, and that it can be explained by the other interrelated preferences that we have come to endorse. In this way, we *authorize* a preference as part of our integral self and reject preferences that do not cohere with it. As mentioned above, an anomalous or alienating preference that is rejected would be considered part of the individual in some way; however, it would not be part of his authentic or integral self. This would then lead to the conclusion that autonomous action requires that an individual act on the basis of preferences that are authorised and cohere, in the way that has been described above, and that are not the product of any form of compulsion or force.

Of course, an objection may be made that the division of a character into two selves, an integral and a secondary or peripheral self is highly problematic. However, Ekstrom is careful to point out that she is not attempting to provide "an ontological thesis" (2005a:152) regarding the existence of some immaterial Cartesian entity. Rather, she is once again drawing upon a popular conception of the understanding of the self, in which such statements would make utter sense. In other words, there is clearly a basis for distinguishing one individual from another in terms of differences in their attitudes, values, preferences and behaviour. More specifically, as individuals, we regard certain preferences as more characteristic of who we 'truly' are than others.

For example, an individual may describe herself as being the kind of person who has a strong aversion to risk-taking endeavours. This could be an integral part of her character, so much so that her friends would know not to include her in plans to sky dive, for example. On Ekstrom's account, this could be further unpacked. We could then ascertain that risk-taking endeavours do not cohere with her other accepted preferences, some of which could be the fact that she is a mother and has a strong preference to be what she considers to be a responsible parent, and a further preference for feeling safe and avoiding danger. There could be a number of other cohering preferences that form an integral part of her character, and, if risk-taking is not supported by this network, then she would reject it as a possible motive for action. On the other hand, if she happens to not have a network of preferences that would either support or reject risk-taking, it would be regarded as peripheral to her character and she would therefore find herself having a lack of strong preferences either way for sky diving, and thus, may feel neutral regarding whether or not she would engage in such an activity.

To provide further strength to her claim, Ekstrom gives three justifications for her argument that those preferences that cohere with one another, along with the evaluative character forming ability, comprise the integral self that gives substance to the notion of self in self-determination, rather than those preferences that fail to cohere or are peripheral to the self. Firstly, the cohering network of preferences imply an integral self because they are *enduring* in character. In other words, preferences that are integral to an individual's character have generally persisted through time, and have thus been recognised as integral to her character with the result that reasons or explanations can be given for them. It is not that they are fixed in perpetuity, but rather, that they are not subject to incessant change. Secondly, as alluded to above, the cohering network of preferences can be *defended* or justified by the individual with sound reasons if they are disputed. In other words, preferences that are part of this integral self would be preferences that an individual would be willing to recognise as his own and he would do so by referring to other preference states that he has which would be able to support and justify them. Even in the case of self-doubt, an individual could come to a position of self-acceptance, and thus, could provide a self-justification if he were to ascertain that a preference that has elicited doubt is congruent with his cohering network of preferences and is thus indicative of his integral self. Ekstrom discusses some examples here.

The first example would be an individual with an unwanted gambling addiction whose preference to gamble would be described as enduring and would thus meet the first requirement for it to be considered part of her integral self (1993:608). However, if someone were to identify her unwanted addictive desire to gamble as part of her core self, if she is an unwilling gambler, she would, in all likelihood, go to great pains to explain to them why they are mistaken and would do so by describing other integral aspects of who she is. She would, however, be willing to defend her preference for the opposite desire, namely that the desire to *not* gamble, be effective in compelling her to act, where acting is refraining from gambling. Thus, the unwilling gambler's preference for compulsive gambling would not meet the second requirement to be considered as part of her integral self, as she would not be willing to accept her destructive preference as part of who she truly is, and thus, to defend it. To the extent that she continued to gamble, thus acting against her core network of preferences or integral self, her autonomy would be compromised in some way.

The second example is a particularly interesting one. We can imagine an individual who purchases a lottery ticket, despite the fact that he has a variety of attitudinal states that do not support doing so. In such a situation, one could say that in purchasing the ticket, and thus acting upon a preference that is not part of his cohering network of preferences, or integral self, his purchase of the ticket

has nevertheless been a free act. He has not been compelled to buy it. Ekstrom posits that while this may be true in terms of "free action" (2005b:59) it is not necessarily true of autonomy which she argues is possibly a "thicker notion" (ibid.) than freedom. Thus, if the individual finds himself acting upon a preference that is anomalous in terms of his cohering network of preferences, it would make sense for a friend to ponder "what has come over him" (Ekstrom, 2005b:59). If the individual experiences a sense of internal discord, this could be "sufficient to undermine [his] autonomy" (ibid.).

Finally, the third justification Ekstrom provides is that a preference that has been authorised by the cohering network of preferences is associated with the integral self, more so than an anomalous preference, if the individual endorses or *approves* of this preference as hers. In other words, it must be a preference that the individual is "comfortable owning" (Ekstrom, 2005b:60). Because such preferences are coherent with her character, and more specifically, her integral self, they would be supported by the network of cohering preferences, and thus, when she acts upon such a preference, she would be said to be "wholeheartedly behind what [she] does…acting…without higher-level reservation" (Ekstrom, 1993:609) If a preference causes internal consternation or anxiety for the individual this is indicative that it may not be supported by her network of cohering preferences, and would thus not be authorised by her.

Thus, if I have a network of cohering preferences that would be described as self-serving, when I find myself with a selfish preference this would elicit no discomfort, as there would be no conflict between such a preference and my network of cohering preferences. I would therefore easily be able to justify such a preference as acceptable to myself. On the other hand, if my network of cohering preferences is predominantly other-centred, a selfish preference would come into conflict with this network, and potentially cause me discomfort and shame, with the result that I would not wish others to know about it and would not seek to defend it. Because Ekstrom's account is content-free with regard to the good, it therefore includes the possibility that one could have a bad character, and thus act autonomously upon preferences that are congruent with this character, but are nevertheless immoral.

Ekstrom posits that one could respond by pointing out that individuals frequently experience a sense of inner discord regarding certain preferences, and this does not always imply that their autonomy has been compromised. However, her point is that while the self comprises the sum total of preferences, beliefs and attitudes, along with the ability to deliberate about these mental

states, the true or integral self would be only those "preferences and acceptances that cohere together" (Ekstrom, 2005b:60), and thus, that discern the individual as an individual. Thus, I may have conflicting preferences, such as a desire to both eat cake because it tastes delicious and to not eat cake because I believe that it will make me fat, and both these conflicting preferences would, of course, be mine, or part of my *self*. However, if one of these preferences coheres more with my other preferences then it could be said to be part of my integral self rather than my peripheral self.

As Ekstrom admits, this specification – that to be considered one's own, and thus autonomous, a preference must cohere with an individual's network of mutually supported preferences – delivers a rigorous theory of autonomy. However, as a rich account of autonomy, I argue that it would have more use-value, particularly in an applied ethics context, than a thin account of autonomy which stipulates the term as referring purely to an undeveloped notion of self-determination. An additional response to Ekstrom's coherence requirement could be that it is akin to a self-reinforcing loop, and is thus an extremely subjective interpretation of autonomy. However, this would be a misguided response as while Ekstrom's interpretation is rigorous, it is not controversial. As Ekstrom points out, there is widespread consensus that "autonomy is to be understood as self-direction, self-command, or self-rule. It is opposed to rule-from-without enslavement, and victimisation" (2005a:155) – and I would add – rule-from-within enslavement and compulsion. Furthermore, as Korsgaard argues, "autonomy is commanding yourself to do what you think it would be a good idea to do, but that in turn depends on who you think you are" (1996:107). In other words, autonomy is, to a certain extent, informed by "practical identity" (Ekstrom, 2005a:161). However, while its content is subjective, its parameters must be defined in objective terms, if it is to carry weight. What Ekstrom has attempted with her theory of autonomy is to support her account of authorisation with a further account that satisfies the second component of Korsgaard's definition. Namely, she has given content to what it means to perceive oneself as a specific entity that acts freely. Ekstrom also argues that when autonomy is investigated in such depth it reveals just how closely related the notions of autonomy and authenticity are (2005a:155).

5.6.5 Autonomy

What remains in giving an account of Ekstrom's theory, is to explain more fully how a cohering preference, and the ensuing action that it supports, receives the status of being considered autonomous. For the purposes of this dissertation, and the way in which I will use Ekstrom's theory

to address the concern for moral autonomy posed by moral bioenhancement, this can be conversely formulated as explaining how her theory is able to deal with threats to autonomy.

It could be said that a preference coheres with my integral self when I come to the conclusion that this preference is more *valuable* for me to have than a competing preference, or, more valuable in conjunction with a preference that together *neutralises* a competing desire, when assessed against my network of cohering preferences or my integral self. A preference that I consider to be more valuable would be one that I would wish to be the basis upon which I act, given my conception of what is true and good. Only when I have come to this conclusion will this preference be personally authorised for me. The notion of personal authorisation is a prerequisite for autonomy as its purpose is to provide an adequate response to conflicting preferences and to work out which preferences are truly, or more, one's own. When I am faced with competing preferences, both preferences will generally be perceived, prima facie, to be true and good, as, if this were not the case then there would be no conflict regarding which preference I wish to be the basis upon which I act. When faced with such a situation of competing preferences that are both considered true and good, a resolution can only be achieved if one of the preferences is *defeated*. This occurs when I come to the conclusion that one of my preferences is unequivocally *more* valuable for me as a preference. As mentioned above, in situations that are more complex, it may be the case that resolution is achieved through adding a neutralising preference to one of my preferences, which then acts together with that preference to defeat the competing preference. The preferences that are rejected will be those that, after having been deliberated upon, have not received personal authorisation, in terms of the fact that I would not wish them to be the basis upon which I act due to their being unsupported by my integral self.

Some examples will be helpful in elucidating the above stipulations. I could be faced with a desire to spend the evening working on revisions for a journal article that is due imminently. This desire is regarded as a valuable one for me to have, as I wish to finish the article to enjoy the feeling of fulfilment of having done so on time, and to have made an active contribution towards my publication record. As I regard all of these factors to be worthwhile outcomes, I am therefore in no doubt that my preference to complete the article is one that I endorse. On the other hand, I may have a good friend who has invited me to dinner that same night, at a restaurant I happen to love. I may have a strong desire to go to dinner with my friend, as I know it will be more enjoyable than sitting alone and working, and it will be valuable to spend time with my friend whom I have not seen for a considerable length of time.

I am thus faced with the decision of which preference I would like to be the one that would be the basis upon which I act. This is a deliberation that will be informed – either explicitly and consciously, or implicitly and unconsciously – by the cohering preferences and attitudes that comprise my integral self[83]. When deciding what I should do, I may then decide that my preference to stay at home and work on the article is a more valuable one for me to have than my desire to go out with my friend, and thus, that I would prefer that this be the desire that is the basis upon which I act. The reason for this could be that while I value friendship, and enjoy the experience of eating at good restaurants, there are other elements of my integral self that may be more closely related to the preference to work on the article, and thus, more supportive of it. Such elements could be that I have a strong sense of ambition, and thus, that I place a high premium upon achieving success in my chosen work. I could also regard having a strong and reliable work ethic as highly favourable, have a distaste for failing to meet my deadlines, a compelling desire for job security, and so on. Thus, the qualities that comprise my integral self are extreme conscientiousness, a strong sense of responsibility, high levels of ambition and a need for security. As mentioned above, I may value friendship, good food and relaxation, but these preferences may be more part of my peripheral character than my integral self. Thus, if I then go ahead and meet my friend at the restaurant, it may elicit feelings of guilt and anxiety, which may cause me to wonder why I have acted out of character and what has come over me in choosing to do so.

The second example would aim to address situations in which I am faced with two competing preferences that are both supported by my integral self. In such a case, I could use a neutralising preference as a means of resolving the issue. In the example discussed above, my integral self could comprise the above elements as well as a strong regard for friendship which would make the decision more challenging for me. I could then decide to both work on my article and see my friend, by suggesting that we meet later in the evening for drinks rather than for a lengthy dinner. I would end up having a late night, and thus some sacrifice would be required on my part, but my preference to spend an entire evening with my friend will have then been neutralised by the addition of rather simply seeing her at the end of the evening, thus enabling me to finish writing the article. In this situation, I am unable to authorise either of the initial preferences as the basis upon which I would act; however, my preference to work on my article *and* see my friend later, but for less time,

[83] In the case where this process occurs in an implicit or unconscious manner, this does not imply mindlessness, as the individual would still be able to provide an explanation for which preference she takes to be more indicative of her integral self, if pressed to do so.

is a preference that I *am* willing to authorise as the basis upon which I will act as it coheres more with my integral self than either preferences do on their own.

Thus, with the above definitions and examples in mind, Ekstrom defines her coherence account of autonomy in the following way:

> An act is autonomous [if and only if] it is nondeviantly *caused* by an uncoercively formed, personally authorized preference. A preference that is personally authorised for an individual has authority for speaking for her, for representing what she truly wants in being well supported by a network of her considered attitudes; it is an attitude with respect to which she is wholehearted. Thus, action on such a preference is self-directed or self-ruled, rather than heteronomous (own emphasis, 2005b:64)[84].

In terms of the problem of authority, also known as the problem of identification, the source of authority that would identify one preference as being more my own than another, and thus, part of my integral self rather than my peripheral or secondary self, would be the process that leads to my authorising it. Thus:

> To identify oneself with some desire is to have a personally authorised preference for that particular desire to be the one that leads one to act, when or if one acts. To identify oneself with some belief is to have an acceptance regarding the content of the belief that is coherent with one's character system. And to identify oneself with some course of action is to perform that act because one has a personally authorised preference in its favour (Ekstrom, 2005b:64)

5.6.6 Alienation

For the purposes of using Ekstrom's account to assess the concern for autonomy posed by moral bioenhancement, the matter that is of particular relevance, is what her theory may tell us of desire-formation and action that would *not* be considered autonomous. One way of understanding a lack of autonomy would be to distinguish between two different kinds of desires. On the one hand, an individual may find himself having a desire that he considers incongruous to whom he considers himself to be. This desire may elicit considerable distress and be rejected by the individual in question as a legitimate basis upon which to act. Such desires may be described as "ego-dystonic" (Ekstrom, 2005a:157), referring to various mental states such as beliefs and attitudes, that an individual possesses, but considers to be at odds with his self-conception. On the other hand, "ego-syntonic" desires would be those desires that one identifies with in some way, either in a strong sense whereby they are personally authorised as preferences, or in a weaker sense in which they are part of one's secondary or peripheral self. On a Frankfurtian account, an ego-dystonic desire would be a desire that is "excluded entirely as an *outlaw*" (1988:170) by the individual in question.

84 By including the term *caused*, Ekstrom is emphasising the fact that a prerequisite of autonomous action is that a preference and an action must be "causally related" (1993:614). She includes this term to avoid the problem of manipulation.

However, just because a desire is perceived by the individual to be an outlaw desire, this does not always indicate that the desire is not a component of the individual's peripheral or integral self. There are different reasons for this. An individual who finds herself faced with an outlaw desire may refuse to recognise such a desire as part of who she is due to having a distorted self-conception, or she may be "in denial" (Ekstrom, 2005a:157) about the fact that this desire is informed by her integral self. There are many examples of individuals who reject their desires as anomalous for various reasons. An individual growing up in a homophobic context and finding himself with a strong attraction to the same sex may reject this desire, at the expense of his authenticity and personal happiness, as a survival mechanism. A woman who has had an extremely religious or puritanical upbringing may reject any strong feelings of sexual desire as inappropriate, and thus, may identify them as incongruous or the product of some malevolent force, rather than recognising that they are, in fact, normal human desires that she has. Another individual, raised in a family that rejects materialism and values human connection, could possess a strong ambition to achieve financial success and reap the material rewards thereof. While this desire could form part of his character or integral self, he could nevertheless reject it as not indicative of who he truly is and label it as superficial due to feelings of guilt and denial associated with his upbringing. In the above examples, the rejection of desires that are, in fact, part of the individuals' integral selves may be explained by the fact that contextual factors are influencing their rejection of their desires. They may feel shame for their desires, or utilise self-denial as a means of maintaining a coherent self-conception.

However, regardless of the reasons for the denial of such desires, if such desires are, in fact, *preferences*, in the sense that Ekstrom stipulates, that cohere with the network of preferences that comprise the integral self of the individual, then they would be desires that are part of the integral self, whether or not this is something that an individual will admit to[85]. Feeling a sense of identification with, or estrangement towards, a desire is not sufficient for categorising a desire as definitively part, or not part, of the integral self. Ekstrom posits that for this reason both personal

[85] This produces a problem that plagues any account of autonomy as self-determination. This problem is discussed by Berlin in his account of negative and positive liberty, where the latter accords with the interpretation of autonomy as self-determination (1969). Positive accounts of liberty lend themselves to exploitation at the hands of tyrannical regimes that support particular ideologies, for the very reason described in the text to which this footnote is attached. If we take the position that an individual is constituted by an integral self and a peripheral self, where the former is regarded as higher and the latter as lower, and combine this with the assumption that individuals may be mistaken regarding the elements that comprise their higher, integral selves, it is a short step to the position that individuals require 'help' in being able to know and act in accordance with their true selves. As Berlin points out "[o]nce I take this view, I am in a position to ignore the actual wishes of men or societies, to bully, oppress, torture them in the name, and on behalf, of their 'real' selves" (1969:133) and their freedom.

and third-person accounts regarding accepted or rejected desires may be mistaken (2005a:158). In the case of the former, individuals may not recognise desires as integral to their self, for the reasons described in the previous paragraph, and in the case of the latter, individuals may be incorrect in their assessment of what is integral to the character of others. In this regard, Ekstrom's theory, and any theory that associates autonomy with the notion of self-determination must act with the presupposition that one knows oneself and others sufficiently to make an accurate assessment.

However, despite the possibility of such cases of denial, for the concept of autonomy as self-determination to be coherent we have to operate with the assumption that individuals generally know what they truly want, and are able to recognise what the core components are that comprise their characters. This point aside, Ekstrom's theory is of specific importance in cases where individuals are led to action by desires that have *not* been authorised by their cohering network of preferences or integral self. In other words, when I find myself acting on a desire that is contrary to what I wish to be the desire that is the basis upon which I act. In such cases, Ekstrom argues that I act in a way that is not autonomous in terms of the fact that my action has not been determined by my *self* (1993:614). This would be because in doing so, I am acting without motivation, I do not have a preference that has informed my action, and thus, my action is not motivated by a conception of what I take to be true and good.

Once again, there are many examples that may be illustrative of such cases. Some such examples were discussed in earlier sections, namely, those cases of individuals who act upon addictions or compulsions that produce desires that are opposed to their preferences. We would describe such unwilling drug-users, gamblers or purchasers of lottery tickets as having acted in a way that is at odds with how they would prefer to act, and thus, as not acting in a self-determined manner. However, it is not only cases of possible psychopathology or addiction in which autonomy would be vulnerable, we can conceive of more quotidian examples that produce threats to autonomy.

One can think of an individual whose integral self is characterised by a particular network of cohering preferences. She has high esteem for values such as integrity, privacy, and respect for others and this manifests as various preferences that she has: to act in a sensitive manner towards others, to always treat others as ends in themselves, to never take pleasure or overt interest in the misfortune of others, and so on. These preferences would be the ones that she desires to be the basis upon which she acts and how she conducts herself. She may then hear about a particularly horrifying case of terrorism, involving the decapitation of a prisoner, and subsequently learn that

there is video evidence on the internet depicting the actual event. This may then elicit a strong desire to watch the video, one which consumes her attention so fully that it causes her to act upon it. This desire produces feelings of inner turmoil and conflict due to the fact that it appears to arrive in a wholly unbidden manner. The individual has never been one who is motivated by morbid desires, she never slows down when driving past motor-vehicle accidents, and she has never enjoyed watching violent depictions, or hearing about violent acts, for that matter. Thus, her desire to watch the video feels entirely 'out of character' or 'at odds with everything she feels she stands for'.

In terms of Ekstrom's criteria for identifying a preference as constitutive of the integral self, the individual's desire is strong, but it is not an enduring one, in the sense that she has not experienced similar desires in the past. Furthermore, the desire is not defensible with regard to the other preferences and attitudinal states that comprise her integral self, and she is most certainly not comfortable having the desire. If pressed to give an explanation for this desire she would either find this extremely difficult or impossible. However, despite all of this, she may feel driven to watch the video, and upon doing so, may experience a feeling of morbid fascination that is compelling in a way that is abhorrent to her. In this situation, by watching the video, the individual has acted upon a desire that she does not have an authorised preference for. Rather, her authorised preference would be that the desire to *not* watch the video be the one that moves her to act, and thus to refrain from watching the video. Thus, there is no motivation on her part that is connected with an evaluated conception of what is true and good, and no support from her cohering network of preferences and attitudinal states that would support watching the video. Her act of watching the video has the feeling of a compulsion for her, and to this extent it is not indicative of her integral self. In this instance, her ability to determine her 'self' has been compromised, and thus, it would be correct to regard her action of watching the video as lacking in autonomy in some way.

Ekstrom's account is thus able to elucidate the source of such feelings of alienation and inner turmoil that are produced by certain desires and actions. Thus,

> a person acts autonomously only when acting from motivation of a certain sort: one that (1) has undergone critical evaluation with respect to his conception of the good, (2) was uncoercively formed, and (3) coheres with his other acceptance and preference states (2005a:154).

While there may be many occasions in which we experience a sense of inner discord, this will generally be informed by the fact that our character, construed in a wider sense, comprises a plethora of preferences and attitudinal states, some of which may be at odds with one another at certain

times. However, an action is only truly autonomous, in the sense of being truly determined by the self, when it is caused by, motivated by or justified with regard to, the integral self.

5.7 An assessment of Ekstrom's coherence theory of autonomy

As mentioned above, Ekstrom's account of autonomy is content-neutral and largely structural. Firstly, it is content-neutral as preferences and the acts that they motivate are determined by the individual's conception of what is true and good, not by some stipulated account of what is good. This leaves room for individuals to form immoral preferences and perform immoral acts in an autonomous manner, for which they would then be held morally responsible. This also accords with the fact that autonomy is not a normative notion; an individual may use their autonomy for moral or immoral ends. Secondly, Ekstrom's theory is a partly structural account of autonomy, as the latter is a product of how a set of relationships between different components are constituted and function together in a specific way. However, it avoids the problem of manipulation that plagues purely structural accounts of autonomy, such as Frankfurt's, due to containing a procedural requirement which introduces historicity into the account. This procedural requirement is evident in the stipulation that a preference must be formed through a deliberative and evaluative process in accordance with the individual's conception of what is true and good, where this conception will have been formed over time. Thus, it is not simply that I have preferences, but the way in which I came to have them that is important on this account. Furthermore, as the ability to deliberate and evaluate one's preferences is deeply informed by "one's moral or psychological identity, one's exercises of this capacity are 'one's own', barring external manipulation via coercive mechanisms" (Ekstrom, 2005b:57). Thus, a further procedural requirement is added that to be considered autonomous, the formation of preferences may not have been the product of manipulation or coercion.

Ekstrom recognises that her approach could be considered to be an overly "rationalistic" (2005b:57) account of autonomy due to its requirement that preferences must have been the product of some form of deliberation or assessment with regard to a conception of the good. However, her response is that individuals do generally seek to understand their motives for acting and attempt to provide self-explanations for preferences and beliefs. In other words, individuals do not generally act in a purely instinctive or mindless manner, they seek to make sense of themselves and their place in the world, and this includes having adequate and sufficient reasons to act. This is evidenced by the fact that when we witness others acting in a seemingly mindless manner or

perceive ourselves to be acting in such a manner, we generally consider this to be a highly undesirable state of affairs.

Mostly importantly however; Ekstrom's account of autonomy manages to avoid the three problems discussed in section 5.4. I have already alluded to the way in which she does so, in my discussion of her theory in section 5.6 and will therefore not repeat this discussion. However, to reiterate, what is most important for the research focus of this dissertation is the way in which her theory manages to avoid the problem of manipulation, which seems to be the problem that most autonomy theories struggle with. She avoids this problem not simply by stipulating that preferences may not have been formed through manipulation or other forms of external coercion to be considered autonomous, as some other theorists do, but rather, by giving an adequate account as to why this would be problematic. In other words, a preference may only be described as autonomous if it can be explained with reference to the integral self. This is because the preferences and attitudinal states that comprise the integral self "are constitutive of the agent" (Stacy Taylor, 2005:15). Thus, if they are manipulated in any way, the concern becomes not simply for the autonomy of the agent, but for the possibility that manipulation could result in an altered, and perhaps entirely different, agent.

We can then assess Ekstrom's account against Dworkin's criteria for an adequate account of autonomy. Firstly, her account easily satisfies the first requirement as it contains no logical inconsistencies, and is not premised upon any controversial or incoherent assumptions regarding autonomy. Secondly, despite the fact that her theory is formulated in a rigorous manner, it does not issue excessively demanding requirements for the possibility of autonomous action. In fact, on her account, understood in the most simplistic or colloquial terms as 'being true to oneself', autonomy is a relatively achievable empirical possibility. Regardless of the moral status of an individual's preferences and the actions that they lead to, so long as she acts upon an evaluated preference that has been authorised with regard to the cohering network of preferences and attitudinal states that comprise her integral self, and, that reflect a conception of what she takes to be true and good, her preference and action would be autonomous. Autonomy in this sense is achievable.

On the other hand, one could argue the opposite and posit that Ekstrom's theory is a too stringent account of autonomy. This is because her theory would posit that in situations in which an individual has felt compelled to act upon an unauthorised preference, he has acted with limited

autonomy. One could then use her theory to argue that if this is the case, such an individual cannot be held morally responsible for his actions[86]. However, one could respond to this claim in similar manner to the way in which Frankfurt has, by providing a theory of moral responsibility that is not predicated on the principle of alternate possibilities in its ascription of moral responsibility (1969)[87]. In other words, Frankfurt argues that even in cases in which no alternatives were available, the individual is nevertheless morally responsible for their actions due to his distinction between freedom of the will and freedom of action. In other words, my autonomy or freedom of will may be compromised to the extent that I have a conflict between my will and my second-order volition, but my freedom to act or not act upon that will remains intact. The question of moral responsibility is, however, an entirely different and extensive area in its own right, and is, therefore, beyond the scope of this chapter which is concerned specifically with presenting an account of what it means to act upon one's own volitions, and the relevance of this for the concern for moral autonomy posed by moral bioenhancement. Furthermore, for my purposes, a stringent account of autonomy is preferable as if a moral bioenhancement intervention could be shown to not impact autonomy on Ekstrom's account, then there is a good chance that it would be an ethically permissible intervention.

Thirdly, Ekstrom's theory is able to provide a clear account of why autonomy is valuable. On her account, the value of autonomy is predicated upon the goodness of being able to form preferences that lead to actions that are a reflection of who one is, and what one takes to be true and good. Because this latter stipulation is built into her account, acting in this way would be considered to be good, for the individual, by definition. Her account does not aim to address the issue of whether the notion is of comparable value to other principles or values that we take to be good; however, due to the fact that it is content-neutral it would be compatible with achieving other goods, if these are associated with the individual's particular conception of what is true and good. As mentioned above, due to the fact that Ekstrom's account is neutral regarding what an individual takes to be true and good, it does not rule out the possibility that an individual may form immoral, but authorised, preferences that lead her to act in an immoral, but nevertheless autonomous, manner. However, this is the case with any content-neutral account of autonomy. Furthermore, while autonomy is determined objectively by the subjective formation of preferences in the stipulated

[86] This would be considered an undesirable outcome, as it runs counter to the strongly held intuition that individuals *should* be morally accountable for their actions, and that claims of compulsion are not satisfactory justifications for evading this accountability

[87] This was discussed in section 4.8.4 of the previous chapter.

manner, the objective assessment of preferences and actions in terms of their moral worth, and thus, the moral responsibility of the individual, would be a separate matter.

Ekstrom's theory would clearly meet Dworkin's fourth criterion as it is fully compatible with most value systems, due to the fact that it is content-neutral. While it could possibly run into trouble with highly communitarian societies in which value is ascribed to the role of interrelational contributions in the formation of attitudinal states, preferences and values, Ekstrom's theory does not rule out the role that one's context plays in forming one's integral self. Rather, she argues that the integral self consists of the individual's *evaluated* preferences and attitudinal states that are informed by his conception of what is true and good. Of course, an individual's conception of what is true and good will have been influenced by societal and cultural factors, amongst others; however, if his preferences and attitudinal states have survived a process of reflection, are enduring, defensible and he is comfortable with them then when he acts upon one of them that is authorised, he will be acting autonomously. The theory also meets the fifth requirement, as it has both practical and normative applicability. In particular, Ekstrom's account may be used to illustrate why external interferences with an individual's ability to act upon their authorised preferences, and thus their conception of the good, would be a negative thing. In fact, I will use her theory in the next section to support a similar argument; namely, that in certain cases, the concern for moral autonomy posed by moral bioenhancement could, itself, be regarded as paternalistic, to the extent that it seeks to thwart an individual's desire to act only upon authorised preferences, rather than unauthorised preferences. This will be fully explained in the second part of this chapter.

In terms of the final requirement, namely, judgemental relevance, Ekstrom's account would also seem to meet this criterion as it does not contain assumptions that are incongruent with generally accepted claims regarding autonomy. Her account looks at the formation of *specific* preferences and the acts that are authorised, or not, by these preferences, and thus, it implies a conception of autonomy that admits of degrees but is also localised, although she does not explicitly state this. In other words, if an individual lacks autonomy regarding her ability to stop smoking, due to the fact that she has an authorised preference that her desire to *not* smoke be the one upon which she acts upon and *not* her desire to smoke, she would lack autonomy in that particular area of her life. Her global autonomy, as such, would not necessarily be impaired, as she may be autonomous with regard to other preferences that are authorised. However, one could either argue that an impact on autonomy in one area of an individual's life erodes their autonomy to a degree, or, that any impact on autonomy produces effects on an individual's global autonomy. The latter conception would be

founded on the view that autonomy is a threshold concept; an individual either possesses it fully, and in all areas, or not at all, whereas the former is congruent with the view that autonomy admits of degrees, an individual may have more or less of it. Ekstrom does not state which view she supports but both conceptions are generally regarded as acceptable.

5.8 Concluding remarks

In this first part of chapter 5, I have presented the theoretical underpinnings that will inform my analysis of moral bioenhancement that will follow in the second part of the chapter. My main aim has been to justify the selection of Ekstrom's coherence theory of autonomy as a suitable means of assessing the concern for moral autonomy posed by moral bioenhancement. I approached this task of justification, firstly, by discussing a hierarchical approach to autonomy, as I posit that the insights that such approaches contain, capture a conception of autonomy that has great relevance for elucidating the problem at hand. The important insight that the hierarchical approach brought to awareness, is that autonomy is not secured simply by being free to do what one desires to do. Firstly, this is because what one desires to do may not only be the product of undue external manipulation, but may also be due to some inner compulsion, phobia or addiction. Secondly, for various reasons, despite having a strong desire to do something, one may reject this desire and wish that one didn't have it, and in particular, not wish to act upon it. However, despite its appeal, the hierarchical approach requires more substance which is evidenced by the fact that it falls foul to several problems, the most serious one being the problem of manipulation. In other words, despite the appeal of such approaches, the conditions they provide for autonomy are not sufficient to indicate *how* or *why* cases of external interference or manipulation would be problematic. Ekstrom's coherence account is able to address the three problems, and in particular, it is able to indicate why external interferences and cases of manipulation would be inimical to autonomy.

On Ekstrom's account, autonomy is achieved when a desire that I have, that motivates me to act, is informed by a preference to act upon that desire, rather than another, where that preference has been authorised by the fact that it coheres with my integral self. My integral self is informed by a network of preferences and attitudinal states that are in coherence with one another, in that each preference may be explained with reference to the others. Furthermore, coherence is achieved through the fact that such preferences are an enduring part of my character, and, because I am both able and prepared to defend them, due to the fact that I am comfortable admitting that these preferences are part of my character. The process by which I come to have this network of cohering

attitudinal states and preferences is informed by my ability to critically reflect upon these states in terms of the value I ascribe to them in being able to realise what I take to be true and good.

While Ekstrom explicitly states that one's ability to reflect upon and evaluate one's preferences, and the actual preferences and attitudinal states that comprise one's integral self, would be regarded as one's own if they haven't been subject to manipulation or coercion, her theory is able to indicate *why* the latter would be problematic. This is due to the fact that the formation of the integral self is a product of the individual's value system. In other words, what I take to be true and valuable has directly informed the kind of person that I am, and most importantly, the kind of person I wish to be. Therefore, an intervention that would be problematic on Ekstrom's account, would be one that resulted in changes to my value system by introducing values that were previously not held, or, an intervention that compromised my ability to deliberate upon such changes. In such cases, the changes to my value system would, in all likelihood, cause me to have new preferences that were previously not held, or that would not have been authorised by myself before the intervention. Conversely, because of the strength of Ekstrom's account of autonomy and its ability to overcome the problem of manipulation, I would argue that any moral bioenhancement intervention that may be shown to pass Ekstrom's conditions for autonomy would potentially be ethically permissible in terms of the concern for autonomy. Furthermore, I will make a stronger argument in the second part of this chapter and posit that in cases where an intervention passes Ekstrom's conditions, there is a strong likelihood that it could increase the autonomy of the individual concerned.

In terms of other general insights regarding autonomy that I take to be relevant for my analysis in the second part of this chapter, I would argue that autonomy should not be framed in terms of a Kantian ability to act upon an entirely rational will that is devoid of emotional interference, as I would argue that this type of autonomy is impossible to achieve practically. Furthermore, arguing that autonomy is only achieved through being able to overcome contingent factors, such as how an individual 'feels' about something or what is personally and idiosyncratically valuable to him – as these factors may interfere with his ability to act from a sense of duty, where the latter is associated with the dictates of the rational will – seems to imply autonomy's opposite.

Rather, I posit that what is at stake in the moral bioenhancement debate is autonomy as moral authenticity. Moral authenticity would refer to an individual's ability to perform morally relevant acts that may be explained with reference to the beliefs, values, preferences and attitudes that form part of her recognisable and enduring character. This conception is very much a personal account

of autonomy as it is informed by the subject's perception of the extent to which she is free to be moved by her own desires, preferences beliefs and attitudes. However, because an individual may consider herself to be freely moved to act upon a desire that she has, even in cases where that desire has been implanted in her through some form of external manipulation or intervention, an account of autonomy must have adequate conditions that equip it to address such situations. Thus, autonomy is not only adjudicated in an internal or subjective manner but must have objective conditions. I would argue that Ekstrom's theory has sufficient conditions, and in the second part of this chapter I will therefore use her theory to analyse the concern for moral autonomy posed by moral bioenhancement.

Chapter 5b – Assessing moral bioenhancement interventions in terms of a coherence theory of autonomy

5.9 Introduction and overview

In chapter 4, after a survey of the literature, a number of moral bioenhancement interventions were identified as potentially inimical to autonomy. They include:

1) Any intervention that is compulsory or covertly administered
2) Interventions that excessively heighten emotional responses, if this were to result in compulsive behaviour (where compulsive behaviour may be understood as acting without any cognitive reflection or reasons for one's action).
3) Interventions that produce hidden identity or personality changes
4) Interventions that *substantially* alter the identity, core beliefs or value system of the individual
5) Interventions that alter the means by which an individual is able to assess and accept any of these changes (this concern is related to the previous concern)
6) Interventions with irreversible effects

However, while the arguments given in the literature as to why these interventions would be problematic are predominantly valid regarding what their concerns attempt to capture, they are somewhat superficial. I would argue that this is because they use a thin interpretation of autonomy or interpret the concept of autonomy, and thus, what is at stake in the debate, as self-evident. For this reason, I have presented what I take to be a comprehensive theoretical foundation in the first part of this chapter that I will now use to assess, on a deeper level, what the above concerns attempt to capture, in order to illustrate more adequately, if and *why* such interventions would be problematic.

In this second part of chapter 5, I will therefore use Ekstrom's coherence theory of autonomy to assess the status of the concern for moral autonomy posed by moral bioenhancement. I will do this firstly, in section 5.10, by utilising her account, along with various insights from the first part of the chapter, to assess the way in which Harris, as the dominant opponent of moral bioenhancement, interprets and applies the notion of autonomy to argue against moral bioenhancement. In section 5.11 I will then use Ekstrom's theory of autonomy to analyse the specific interventions and outcomes of moral bioenhancement that would be the most likely to impact upon autonomy. Finally, in section 5.12 I will argue that in a certain type of case, moral bioenhancement could

produce an increase in the degree of autonomy experienced by individuals, and therefore, conversely, that preventing individuals from undergoing moral bioenhancement in such cases would be a violation of their autonomy in the Ekstromian sense and would thus be paternalistic.

5.10 An assessment of Harris' argument

Harris raises many valuable and relevant concerns in the moral bioenhancement debate. I would argue, however, that the most important issue that he has raised pertains to the potential consequences of excessively heightening emotional response. Generally, when an individual is faced with a morally relevant decision for which competing options exist, she will make her decision based upon composite cognitive, contextual and affective factors, along with the value system that she endorses which will have been informed by the afore-mentioned factors along with the sum total of her life experiences. This conglomeration of factors will have existed in a particular structural relationship, which will have informed the moral decisions she has, heretofore, made. If moral bioenhancement excessively heightened one component of this structure, namely, the affective component, producing changes that impacted the individual's ability to assess her changed motivations, this would give cause for concern regarding the authenticity of her ensuing decisions. In other words, we could question the extent to which her ensuing moral decisions and behaviour were actually *hers*. The response to this concern has typically been to point out that the proponents of moral bioenhancement are not suggesting that emotionally relevant dispositions such as empathy, be *excessively* heightened. However, despite the fact that they would not be aiming at such radical enhancement, the concern is nevertheless a valid one. The above-mentioned structure is a complex system with interrelated components, and, changes in such systems may produce unforeseen consequences with large-scale effects. This is therefore a concern that merits being taken seriously and will be discussed further in section 5.11.3.

In terms of a point that is related to this concern, what the investigation of hierarchical accounts of autonomy has illustrated is that in cases in which there is congruence between an individual's will and his second-order volitions, this does not rule out the possibility that such congruence is the product of some form of external manipulation. It is for this very reason, that hierarchical accounts are vulnerable to the problem of manipulation; congruence is not sufficient for autonomy. Therefore, as Harris has correctly pointed out, but in a different formulation, in cases of moral bioenhancement where an individual accepts her new will or motivation to act, we cannot assume that her autonomy remains intact simply because of this acceptance or congruence between the two levels. This is because her state of acceptance may, itself, have been a product of the intervention.

Thus, if a moral bioenhancement intervention were to heighten affective components to the extent that it altered second-order volitions, thus impacting upon the individual's ability to assess her will or motivations by means of judgements and preferences that are *truly hers*, in the stipulated sense, then this would be ethically unacceptable in terms of the concern for autonomy. However, I would argue that not all interventions currently discussed in the literature would necessarily do this. I will address this matter in the course of this chapter.

The above issues aside, there are entirely different points that may be made regarding Harris' position. His argument that moral bioenhancement would be inimical to moral autonomy is wholly informed by his interpretation of morality and its requirements. He pays considerable attention to the necessary and sufficient conditions of morality, and, as such, his interpretation can strictly be identified as a concern for the autonomy of morality itself, rather than a concern for the personal autonomy of individuals to be moral, although the two issues are, of course, not unrelated. However, while the universal human capacity for morality may be regarded as an intrinsically good thing, the same argument may be made for the individual capacity to determine one's own life-course in accordance with one's conception of the good. Problems arise, however, when the two goods come into conflict with each another, which is essentially the crux of the ethical concern for moral bioenhancement. In other words, reframed in this way, the concern is that moral bioenhancement, due to the fact that it will impact on human moral autonomy as such, should not be permitted as this will signify an erosion or destruction of something that is a valuable characteristic of humanity as a whole. Therefore, the conclusion is that regardless of whether or not individuals would wish to exercise the choice to morally bioenhance themselves, they should not be permitted to do so. Many of the arguments in the literature, and I would argue Harris' argument in particular, operate with this implicit assumption, therefore, it would be more accurate to describe their concern as one for human morality as such, or for universal human moral autonomy.

Another problem with the Harris line of argumentation is that, by framing his argument in terms of the stipulation that moral autonomy requires the *freedom to fall*, he is implicitly employing a substantive account of autonomy. This is because his conception of autonomy has a condition – albeit a subtle one – regarding what an individual may take to be true and good. The matter is difficult to explicate, as, in the case of an investigation of moral bioenhancement, we are not simply assessing the autonomy of general preferences and actions, we are narrowing the focus to assess the autonomy of *moral* preferences and *moral* acts. With this latter focus, the matter risks becoming

clouded by normative implications rather than remaining focused on the conditions of autonomy. However, the same conditions should apply when judging preferences and acts in terms of their autonomy; this should not change because of the content of these preferences and acts. In other words, the moral status of an action is a separate matter to the autonomy of an action.

Harris' approach does not produce problems for cases of compulsory moral bioenhancement, nor for cases in which there is genuine congruence between an individual's authorised preferences and their actions. I would argue that this is because moral bioenhancement in the former instance would be an unacceptable violation of negative liberty, and therefore, would be ethically impermissible on any account. Thus, Harris' conception of moral autonomy would not be required to add impetus to the argument or to do any additional 'work' in this regard. In the latter instance, on the grounds of autonomy as self-determination, and as something intrinsically valuable for individuals, biologically altering a desire that an individual has an authorised preference for, regardless of the moral status of either the preference or the desire, would be morally problematic in terms of its impact on his autonomy. In other words, if there is congruence between an individual's authorised moral preferences and his actions, the individual would presumably not experience any inner conflict in this regard, and would thus see no need for moral bioenhancement. In such a case, forcing him to undergo moral bioenhancement would be an *undue* violation, not only of his freedom, but also of the dictates of personal autonomy, regardless of whether his authorised preferences and actions were considered by others to be morally praiseworthy or immoral in character[88]. However, once again, to illustrate why this would be problematic, we do not need to utilise Harris' argument that morality requires the freedom to fall.

In a particular type of case, however, Harris' argument that moral autonomy, or morality itself, requires the freedom to fall, and thus, that moral bioenhancement – to the extent that it lessens the possibility of choosing to fall – is absolutely wrong, disregards the possibility that an individual may have a desire that *not* falling – where this is associated with avoiding doing something that she perceives to be morally problematic or abhorrent – might be the preference upon which she wishes to act. In this regard, Harris is adding subtle content to the notion of autonomy by ruling out a possible conception of the good that an individual may have. This would be the specific preference that *not* being able to act upon particular desires that she might have, may, for that

[88] I emphasise the term 'undue' because an individual would, of course, be held morally responsible for his preferences and actions, to the extent that if he illegally violated the freedom of others, resulting in his incarceration, then his freedom would be curtailed. However, as discussed in chapter 4, I agree with the prevalent view in the literature that the loss of freedom of action from imprisonment is different in kind to the loss of freedom that would occur from being forced to undergo a biological intervention such as moral bioenhancement.

individual, be part of her conception of the good[89]. In this regard, Harris' account falls foul to the criticisms that plague substantive accounts of autonomy in general. In other words, by stipulating a particular conception of the good as a determinant for moral autonomy, his account would be guilty of subtle paternalism and logical inconsistences.

By this, I am not arguing for the absolute value of autonomy as self-determination. Regarding the ability to self-determine as valuable in an absolute sense is highly problematic, I would argue, as this would support the ability to perform self-determined actions that are both good and evil. In this regard, I agree with Persson and Savulescu that the world would be preferable without certain decidedly immoral, but self-determined, acts. Thus, in certain cases, a strong consequentialist argument may be made that an intervention that prevents individuals from performing certain acts would be justifiable regardless of any loss of autonomy. However, this is not a line of argumentation that I have chosen to investigate in this dissertation. Rather, I have specifically attempted to investigate whether moral bioenhancement, as I defined it at the end of chapter 2, would be likely to impact moral autonomy in the way that critics fear it would. In this regard, I am positing that if the critics are lodging their arguments against moral bioenhancement on the basis of the concern that it will negatively impact autonomy and we agree that autonomy implies self-determination, then, *by definition*, it cannot be logically coherent to also agree that it is acceptable to place restrictions on the content of self-determination in order to protect autonomy as self-determination.

The incoherence of this position indicates that although the debate is framed in terms of a concern for autonomy, the concern is rather for human morality, as such, which is regarded as intrinsically valuable and which should therefore not be altered. In this regard, I am not arguing that autonomy theories are not subject to restrictions in terms of procedural and structural requirements, as this is a condition for something to be considered a theory rather than a collection of claims. Rather, I am arguing that as soon as there are restrictions placed upon the content of self-determination, particularly in terms of moral content, then we can question whether it is still the concept of self-determination that we are referring to. Of course, I fully acknowledge that with the right to self-determination, if there is such a thing, comes the obligation or duty to be held morally responsible

[89] Examples of pathological desires that may have been rejected by the individual are, of course, easy to provide. However, in terms of moral bioenhancement, rather than the treatment of pathological functioning, an example would be an individual who has problematised her levels of empathy, where this leads her to have desires that she regards as causing her to act in a selfish or immoral manner. If she rejects such desires on the basis of who she takes herself to be and what she regards as good, and she is able to undergo an intervention that would increase her levels of empathy, it would be problematic to prevent this on the grounds of protecting her autonomy.

for one's actions. However, I take this to be an area of investigation that is distinct from the issue regarding the conditions for autonomy, and one that therefore, requires its own arguments.

5.11 Interventions that would violate freedom and/or autonomy

5.11.1 Compulsory Interventions

As mentioned above, and discussed in chapter 3, the most ethically problematic type of moral bioenhancement intervention would be any intervention that is forced on individuals against their will or without their knowledge. In other words, any intervention that is compulsory or covertly administered. Such interventions would clearly violate both freedom and autonomy due to considering neither. As discussed in previous chapters, in order to justify compulsory interventions, one would have to do so on utilitarian grounds. One could do so by arguing that the value of autonomy is overstated, that there are some 'goods' that we value more in terms of their utility, and thus, that a trade-off between autonomy and other goods is worthwhile. This is the approach that Persson and Savulescu take, although they formulate their argument by framing the avoidance of harm as a good. Such an approach requires providing strong grounds for the claim that the harm associated with compulsory moral bioenhancement would avert a greater harm, such as existential risk. However, even in such cases, the response may be that the former would be a greater harm than the latter. In this regard, Harris has explicitly posited that he "like so many others would not wish to sacrifice freedom for survival" (2016. 74-75). I will return to this point in the conclusion of this dissertation. These kinds of arguments, which have characterised the moral bioenhancement debate, have reached a point of deadlock as they represent clashes between opposing value systems that are fundamentally irresolvable. If one agrees that existential risk is a distinct possibility, then, fundamentally, the issue becomes what one would and would not be willing to advocate and do, in order to ensure the survival of the human species.

However, these points aside, it is generally accepted in the literature that compulsory moral bioenhancement would be an undue violation of negative liberty or freedom, and thus, any argument that is utilised to illustrate why any such violations are morally indefensible could be used to argue the same for compulsory moral bioenhancement. Furthermore, in terms of Ekstrom's account, compulsory moral bioenhancement would clearly violate her conditions for autonomy. This is because it would essentially amount to a form of unwanted external manipulation of desires and possibly preferences, the possibility of which her theory explicitly seeks to exclude. However, while in the case of an argument against compulsory moral bioenhancement, an account of personal

autonomy would not be required to do any extra 'work' that could not be done by arguments that are able to illustrate why undue violations of freedom are themselves wrong, we could still use Ekstrom's theory to illustrate further, *why* moral bioenhancement would be a violation of autonomy both in cases of compulsory moral bioenhancement and in *certain* voluntary cases.

5.11.2 Creating unsupported preferences

Assuming compulsory moral bioenhancement to be an undue violation of liberty, and thus, to be morally indefensible, we can then turn to other possibly problematic interventions. This investigation would include only voluntary moral bioenhancement interventions, thus, there would be no violation of freedom or negative liberty in such cases[90]. In other words, the concern would be that despite being voluntary, moral bioenhancement may still impact upon personal autonomy or the ability to truly determine the self. The threats to autonomy in this realm would admit of degrees in that there would some outcomes that would more severely impact autonomy than others. Using Ekstrom's requirements, the first level of threat to autonomy would be the introduction of anomalous desires or the heightening of weak desires. In other words, such desires could be desires that the individual had previously not possessed, or, they could be desires that had been part of the peripheral self, which, due to being strengthened via a moral bioenhancement intervention, would now claim the attention of the individual, thus becoming preferences. On Frankfurt's account, this would be a process whereby moral bioenhancement caused an existing first-order desire to become the will of the individual or created a new first-order desire that became the will of the individual.

To illustrate by way of an example, we can imagine an individual who focuses her attention on matters that are deemed to be of relevance in light of her cohering network of preferences and attitudinal states. When encountering poor or needy individuals, she may experience a fleeting sense of empathy for their plight but not have a strong preference in terms of the way in which she would wish to be moved to act in this regard. Rather, her existing levels of empathy could manifest as a concern for the welfare of animals rather than human beings, and thus, as a desire to rather help the former than the latter. Thus, empathy for the poor would be part of her peripheral self: she may think about it at times and consider it to be unfortunate, but not consider it important enough to motivate her to act in any way.

[90] However, the arguments in this section would also serve as a means of explaining, by way of an argument from autonomy rather than an argument from freedom, why compulsory moral bioenhancement would be morally problematic. In other words, the argument would be that it is wrong on account of freedom-related arguments *and* autonomy-related arguments.

On the other hand, we can imagine an individual who has a cohering network of preferences and attitudinal states that include among others, a belief in the goodness of being highly ambitious and self-sufficient through hard work as well as preferences to be a person who is extremely successful, professional and clear-minded. This individual may have deliberated upon the plight of the poor and come to the conclusion that one should not give 'handouts' to such individuals due to the possibility that it may increase their dependence on such acts of charity occurring, and thus, make it less likely that they could achieve self-sufficiency. This belief may be informed by the fact that she has low levels of empathy in general. Regardless of the moral status of her preferences and beliefs, these attitudinal states would be a reflection of what she takes to be true and good, so that she would wish that the preference upon which she is moved to act be a preference to not disempower the poor by making them more dependent. In this second case, we would say that there is congruence between her preferences and desires regarding this matter. When she refrains from giving money to the poor she is acting upon a preference that has been authorised by her integral self.

In terms of the first example, if the individual were to then choose to undergo an intervention – due perhaps to being offered an incentive to do so, or for some other reason – that strengthened her empathy levels, in the case where she once only had fleeting concerns for the plight of the poor, she may now have a *stronger* desire to help them, one which is now sufficiently strong to move her to act upon it[91]. In such a case where a weak, but pre-existing, desire has been strengthened, I would argue that while this would not be wholly unproblematic, it would be less so than in the case of the second individual who had no such desire before the interventions. This is because the individual in the first example already had an *other-focused* desire to assist those in need which was motivated by the presence of existing levels of empathy. It just so happened that her empathy moved her to be concerned for, and to act upon, the plight of animals rather than human beings. After the intervention, her empathy would now have been strengthened sufficiently to include both human beings and animals as targets of her concern. Thus, it would not necessarily be the content of her attitudinal states that would have been changed; rather, it would be the focus or scope of her attitudinal states that would have been widened.

[91] In the case of an individual *voluntarily agreeing* to undergo moral bioenhancement, despite the fact that she has congruence between her desires and preferences, and thus, experiences no sense of conflict, we could argue that this implies that on some level there isn't really congruence between the two. In the case of the first example, the individual could have experienced some form of guilt regarding her attitude towards the poor and felt a desire to change this attitude. This would then indicate that her careless attitude towards the poor is not genuinely part of her integral self and that she has problematised this attitude in some way because it is odds with her integral self. I will discuss this matter further in section 5.7.3

On the other hand, in the second case, the individual may now have an entirely *new* desire to help, that could have replaced her old desire and her belief that would not regard such assistance as truly helping. To the extent that the strength of her new desire caused her to now have a new preference to assist the poor, and in particular, that this preference be the one upon which she would want to act, this would have impacted her autonomy in some way. In other words, the new preference would not have been authorised by way of its coherence with her network of preferences and attitudinal states, that, in turn, is informed by her conception of what is true and good; it would be an anomalous or 'rogue' preference. In fact, to the extent that the new preference would be at odds with her integral self, this would be a serious impact upon her autonomy. Furthermore, to assimilate such a change into her integral self, she would have to come up with post hoc rationalisations to do so. These reasons would not be an accurate reflection of what she takes to be true and good, and thus, would be inauthentic. This second category, therefore, refers to any intervention that creates a new desire which is strong enough to become a preference and is sufficiently compelling to cause action, despite not having been authorised by the integral self in the stipulated manner. Interventions that strengthen an existing but weak desire to the extent that it becomes a new preference that is now sufficiently compelling to cause action would require attention, but I would argue that the former would be a far more serious impact on autonomy as self-determination than the latter.

5.11.3 Changes to the integral self

There is, however, a potential concern that is even more serious in its implications for autonomy than the above-mentioned, which I will now investigate by means of the second example discussed in the previous section. As mentioned above, in the second example, the individual's integral self is characterised by a cohering network of preferences and attitudinal states that include among others, a belief in the goodness of being highly ambitious and self-sufficient through hard work, as well as the preference to be a person who is extremely successful, professional and motivated by rational rather than emotive factors. She may also be an individual who has lower levels of empathy, and thus, she may have no definitive desires or preferences to help others. To formulate this in Ekstrom's parlance, a preference to help others is simply not regarded as one that is valuable for her to have in light of what she takes to be true and good.

Imagine then, that this individual decides to undergo an intervention that enhances her empathy levels. Her decision to do so may have been informed by a strong incentive that appeals in some way to her ambitious nature. The concern that this third category attempts to capture would be if

the increase in her empathy levels were to produce significant changes to the above-mentioned cohering network of preferences and attitudinal states. In other words, if she came to realise, due to new feelings of empathy and sensitivity, that ambition and personal success were relatively trivial goals, that self-sufficiency is meaningless in the face of the suffering of others, and that seemingly rational clear-mindedness is cold-hearted, such changes would have substantially altered her integral self. While it could be argued that the changes produced by the invention in this case were decidedly positive and that post-enhancement her new dispositions were morally preferable, or even that she is now a much better and nicer person in general, I would argue that this response misses the point of what is actually being addressed here which is the question of whether or not all moral bioenhancement interventions would produce impacts on autonomy. I will return to this point below.

On Ekstrom's account, the most serious concern posed by moral bioenhancement would be that in strengthening a particular disposition, such as empathy, this may produce more far-reaching effects. As discussed in section 2.4.2 of chapter 2, a number of critics of moral bioenhancement have discussed the dangers posed by moral bioenhancement due to the nature of human moral psychology which exhibits "ontogenetic and neuropsychological" (Zarpentine, 2013:145) complexity, and, is further complicated by the fact that cognition is domain-general, rather than located in a specific area of the brain (Young & Duncan, 2012:1). Agar lodges a similar concern that isolating and strengthening one disposition may produce imbalances in other areas leading us to support moral judgements that we have previously not approved of (2015a:344), or, it may impact upon the mental flexibility required to balance other important factors that are integral to morally appropriate outcomes (Barilan, 2015:79) The example discussed above happened to be a case with an outcome that would be regarded as positive by most, however, there is no certainty that all cases would produce such outcomes.

The above possible outcomes would be associated with the concerns discussed in chapter 4a, namely, impacts upon personal identity. In the parlance of Ekstrom's theory, this would potentially be an impact upon the individual's ability to assess his desires in terms of the cohering network of preferences that comprise his integral self and his conception of the good, due to a sufficient number of changes in this cohering network of preferences itself. In other words, the enhancement of empathy, for example, rather than simply strengthening a specific desire, and thus, a particular preference, could alter the moral cognition and moral psychology of the individual in general. Thus, the concern is that change in one area could alter the preference structure as a whole, and

this would be the same as altering the integral self. On the continuum of potential impacts to autonomy this would be the most undesirable outcome as major changes to the integral self could produce a qualitatively different self, thereby eradicating the means for moral self-determination entirely.

As discussed in chapter 4a, certain neurological interventions, could pose distinct risks to personal identity. In particular, there is empirical evidence of individuals exhibiting strong identity changes after having received deep brain stimulation for various movement and affective disorders. In the case of what are regarded as extreme personality changes, the individual may be described by those close to him as being an entirely different person, in a qualitative sense, after such interventions. While there is not absolute consensus regarding what is implied by the kind of identity that is potentially vulnerable in such cases, qualitative identity has been identified in the literature as the most likely contender. This type of identity is descriptive and aspirational in that it is informed by how individuals conceive themselves – the self that they believe themselves to be – as well as the selves they aspire to be. More specifically, narrative identity has also been identified as the particular type of qualitative identity that would be vulnerable to impacts from moral bioenhancement. This account of identity includes not only an individual's self-conception, but also his interpretations of the events of his life, which would include the way in which he assimilates any changes to his self-conception, his relationship with others, as well as what he values and identifies with.

The similarities between these descriptions of identity and the notion of Ekstrom's integral self, comprising the enduring preferences and attitudinal states that form a cohering network and are informed by the individual's conception of what is true and good, are obvious. In fact, she uses the terms moral or "psychological identity" (Ekstrom, 2005a:154; 2005b:55) a number of times, to capture what is implied by the integral self. While the notion of an integral self is not unproblematic, Ekstrom's psychological or moral conception of the self largely avoids metaphysical claims and is premised on the first-person experience of a distinct self that is the locus around which desires, preferences and other attitudinal and belief states can be understood, and most importantly, explained or justified. In this regard, I would argue that personal identity, as explicated in chapter 4a, and Ekstrom's integral self are one and the same thing.

Ekstrom's theory also provides us with a way of responding to Douglas' argument in section 4.2.1 of chapter 4a. Douglas has responded to the concern for identity by arguing that even if moral

bioenhancement did produce identity changes, in the case of negative aspects of an individual's identity, such as certain counter-moral emotions, if such emotions had a negative impact on the individual, change in this regard would be a decidedly positive outcome. Furthermore, one could add that identity changes in the case of those individuals whose identities contain obviously negative dispositions, such as certain counter-moral emotions, would be a good thing for society, in so far as the latter is affected by the consequences of such counter-moral emotions.

This is an obvious and powerful response to the concern for identity as it is difficult to understand the value of holding onto desires and preferences – or, an identity for that matter – that are ill-serving for the individual or determinantal for others, in terms of negative societal impacts they may produce. However, while this may be true for cases in which there is a lack of congruence between desires and cohering preferences or attitudinal states that produce inner conflict and distress for the individual, if autonomy, understood as the ability to determine one's self, is truly regarded as a desirable good then this argument would not hold for cases in which there *is* congruence between desires and preferences. In other words, while an individual may possess a particular counter-moral desire, if she authorises this desire to be the one upon which she acts, due to its coherence with her network of preferences or integral self, then this signifies her self-determination. Formulated slightly differently, if the preferences that form part of an individual's cohering network, and thus her integral self, are enduring and she is comfortable possessing and defending them to herself and others, as well as able to explain them with reference to her other preferences, beliefs and attitudes, then these preferences and the actions that they authorise, as well as her integral self or identity are self-determined[92].

Nevertheless, one could persist with this line of argumentation and provide numerous claims as to why it would be desirable to change those aspects of individuals' identities, or their identities in their entirety, that we do not approve of. However, in the case of those individuals that truly accept their desires and identities as reflective of their conception of what is true and good, regardless of the moral status thereof, all such arguments would be set to collide with the dictates of autonomy[93].

[92] Because the discussion in my dissertation engages with the possible *enhancement* of morally relevant dispositions, such as empathy, to the levels displayed by moral exemplars within society, rather than the *treatment* of pathological dispositions or functioning, it must be remembered that when I refer to an intervention that disrupts a pre-existing congruence between a counter-moral desire and an authorised preference as problematic from the point of a respect for autonomy, I am not condoning such congruence that would support criminal or immoral actions that would be considered pathological. There are arguments that can be made regarding why, in cases of criminal and pathological congruence, a 'corrective treatment' could be a good thing, all things considered. This is, however, not the area that I am focusing on in my investigation.

[93] Once again, I wish to emphasise the point that I am specifically investigating the status of the concern for autonomy as discussed in the literature. One could provide a powerful consequentialist argument as to why it would be a

In addition, where there is a strong regard for self-determination as something intrinsically valuable, this has additional value in serving to act as a protective mechanism against the possibility of moral eugenics agendas that would seek to identify which identities are acceptable and which are not, and should therefore be altered. This would be of particular relevance in cases where those driving a moral eugenics agenda justify their aims with ideologies that are erroneous or morally problematic in nature. It must always be remembered that the interpretation of what would be considered an unacceptable or acceptable identity is a product of cultural and historical context to a large extent. It is not only identities containing components regarded as definitively immoral – such as those consisting of preferences to rape, abuse, exploit and murder – that are regarded as problematic. In the past, there were identities, considered to be problematic, that were characterised by preferences that are now considered not only legitimate but morally valuable. A desire to be in a loving relationship with someone of the same sex; a desire, as a woman, to vote, work, and achieve self-sufficiency and power; and a desire to be considered politically and morally equal, regardless of one's race, would be just three examples of identities that would have been considered problematic in the past, and, in certain areas of the world, are still regarded as highly problematic. The point here is that in addition to the possibility of biologically altering identities that would be regarded as uncontroversially problematic, there is also enormous potential for exploitation and abuse.

The discussion in this section, as based upon certain insights from chapter 2 and chapter 4a, and interpreted in terms of Ekstrom's theory of autonomy, indicates that moral bioenhancement interventions that would produce major impacts to identity, where there is congruence between identity, desires and behaviour would be potentially inimical to autonomy as self-determination. This would be the case even if an individual, for whatever reason, has agreed to voluntarily undergo a moral bioenhancement intervention. In other words, the discussion in this section provides us with more substance as to *why* the concern for identity, as presented in the literature, would not be unfounded.

worthwhile trade-off to impact upon the autonomy of individuals who have congruence between morally problematic desires and their identities by forcing such individuals to undergo moral bioenhancement. This, of course, is the argument that Persson and Savulescu make. However, such an argument will hold no weight with thinkers, such as Harris, who respond from a non-consequentialist position and would regard this as an undue violation of autonomy despite any positive consequences it would produce.

5.12 Interventions that would not violate autonomy

While the interventions discussed in sections 5.11.2 and 5.11.3 represent ever-increasing degrees of magnitude in terms of their potential impact upon autonomy, Ekstrom's theory may also be used to argue for a special kind of case in which moral bioenhancement interventions could be regarded as not only morally permissible, but as morally desirable in terms of their impact upon autonomy. This would be in cases characterised by a lack of congruence between an individual's desires and preferences, and, more specifically, in cases in which an individual habitually and compulsively acts upon desires that have not been authorised, in the sense in which Ekstrom uses the term.

However, it must be pointed out that a lack of congruence between desires and preferences is not necessarily, on its own, an indication of inauthenticity or compromised autonomy as we are frequently faced with such instances. Frankfurt's theory eloquently elucidates the insight, that to be human is to experience a degree of conflict between one's instinctual feelings, desires, beliefs and attitudes and the feelings, desires, beliefs and attitudes one would rather have. However, when an individual is continually faced with a desire of such an intensity that she frequently acts upon it, or feels in danger of acting upon it, despite the fact that it is not an authorised preference, and thus, would not be the desire upon which she would wish to act, then, if there is a safe and available intervention and the individual freely chooses to undergo such an intervention, I would argue that this would be morally permissible. In such a case, and on Ekstrom's account, if an individual is frequently acting upon an unauthorised preference, then her self-determination, and thus her autonomy, may be said to be compromised. Therefore, to the extent that a moral bioenhancement intervention would result in congruence between her preferences and desires, by having changed her desire rather than an authorised preference, it would arguably strengthen her ability to act in a self-determined manner. Conversely, refusing such an intervention to her, on the grounds that it is inimical to her autonomy, would be a double affront as it would not only maintain her state of compromised autonomy but would do so by using a paternalistic justification of protecting her moral autonomy as well as – in the case of Harris' line of argumentation – protecting the autonomy of morality in the more universal sense.

Some examples illustrating cases in which there is conflict between desires and preferences have already been discussed in the first part of this chapter. Here, I am referring specifically to cases in which individuals find themselves acting regularly on an unauthorised preference, when their preference would rather be that an opposing desire be the one upon which they act. Some of the examples mentioned were cases of addiction and compulsion, such as unwilling drug-users and

gamblers who feel determined by their compulsive desire to engage in the very habit that destroys their ability to determine their own lives. In such cases, their preference is that their desire to not use drugs or gamble should be the one upon which they act, rather than their desire to use drugs or gamble. However, it is not only in cases of pathological behaviour, addiction and compulsion that individuals could experience an impact upon their ability to self-determine. The category of moral bioenhancement interventions that could produce increases in autonomy may also be extended to the moral realm where individuals find themselves regularly acting in a way that they perceive to be morally problematic or in a way that is in conflict with how they would prefer to act, and which therefore produces a sense of conflict and inner turmoil.

The problematic aspects of human moral psychology that Persson and Savulescu have identified as a means of justifying the project of moral bioenhancement would be examples of dispositions that, if biomedically altered, could potentially produce incongruences between desires and preferences (2012:3-4). Of course, in terms of a respect for a content-neutral account of autonomy, the identification of those desires that would be considered to be potential contenders for moral bioenhancement would be left up to the individual in question. However, any dispositions that have been identified would, of course, have to be amenable to biological alteration. In addition, such interventions could also only be ethically justified *on the grounds of enlarging an individual's sense of autonomy* to the extent that his behaviour is distressing for him due to its conflict with a preference to act upon an opposing desire where that preference is authorised by his integral self. The major proponents of moral bioenhancement, such as Persson and Savulescu and Douglas, frame the problem in terms of the bioenhancement of morally relevant *dispositions* that may also be understood as morally relevant *emotions*. Ekstrom's theory is nevertheless able to account for the role of such dispositions. On her account, a disposition or emotion would inform, or be the basis of, a particular desire to act, or not act, in a specific manner. In other words, as the proponents formulate it, a disposition would *motivate* a desire to act in a particular manner.

To illustrate by way of two examples, an individual could have a desire to cheat in an exam that is informed by the fact that he has a lazy disposition. His laziness would be both a cause, and thus, an explanation for his desire to cheat. His network of cohering preferences may include, among others, a preference to succeed in life with minimal effort, a belief that cunning attributes rather than educational qualifications are important, and a preference to pursue activities that are enjoyable rather than activities that are simply a means to an end. Thus, the individual's desire to cheat would be authorised by his network of cohering preferences and his conception of what is

true and good. He would not problematise his behaviour, and would therefore see no need to change it. While his behaviour would be clearly morally problematic, it would, nevertheless, be autonomous on Ekstrom's account. Therefore, one could not avoid the fact that if he underwent a moral bioenhancement intervention, a cause for concern regarding the authenticity of any behavioural improvements after such an intervention would not be unfounded. On the other hand, another individual may have a desire to regularly cheat in exams due to an anxious disposition and a lack of confidence in her ability to pass without doing so. However, she could also have a preference to not cheat in exams as she may have related preferences such as a preference to pass on her own merits as well as a preference to be honest and act in a manner characterised by integrity. To the extent that her anxiety motivates her self-doubt and compels her to then cheat, thus acting upon her desire to cheat and against her preference to rather desire not to cheat and pass her exam with integrity and in an honest manner, she would be acting in a manner that is product of compulsion rather than self-determination. In this case, an enhancement that would counteract her desire to cheat could be regarded as also producing an increase in her autonomy[94].

An example of a problematic, morally relevant behaviour that Persson and Savulescu have identified, which has been discussed at length in previous chapters, is our lack of empathy for those not closely connected to us which manifests as selfishness and an unwillingness to make small sacrifices that would produce major collective benefits for those living in extreme poverty. This unwillingness to make relatively small sacrifices also has implications for materialistic consumption which, in turn, negatively impacts the environment and exacerbates climate change. A heightened sense of empathy would enable individuals to consider the perspective and position of others and take these matters into consideration when acting. To the extent that an individual is distressed by the low levels of empathy she feels for others, so that she acts in a manner that she perceives to be selfish and wishes that she would not act in such a way, she would be a contender for a moral bioenhancement intervention to increase her empathy levels. In cases in which an

[94] Regarding this example, one could also construct an argument that is informed by the same terms that Harris utilises. In other words, one could argue that an individual, such as the one in the example, could be in a state – pre-enhancement – that is seemingly akin to the state that Harris fears individuals would be in after a moral bioenhancement! In other words, Harris is concerned that moral bioenhancement could excessively heighten emotional responses, thus producing compulsive behaviour. One could then respond to him by pointing out that in certain cases individuals are already acting in a seemingly compulsive manner, which may be informed by a pre-existing level of emotional response that is excessive. Therefore, if a moral bioenhancement were able to safely and effectively address this, in such cases, it would produce a result that is the opposite of that which is feared by Harris. However, one could respond to such examples by pointing out that when behaviour is motivated by an excessive emotional response, this indicates some form of pathology, and it is therefore not correct to regard the solution as an *enhancement*; rather, what is required is *treatment*. In this regard, the difficulty of providing clear cut examples of cases whose solution would be regarded as an enhancement rather than a treatment, indicates the tenuous nature of the treatment/enhancement distinction, as alluded to in chapter 2.

individual experiences distress regarding her behaviour, to the extent that she wishes to utilise an intervention to alter it, this would indicate that the desire that motivates her action, when she acts selfishly, is at odds with her integral self, as well as the likelihood that her preference to not act in such a manner is a genuine one. In such cases, I would argue that a moral bioenhancement would enlarge, rather than compromise, the individual's autonomy.

Douglas' identification of counter-moral emotions as possible targets of moral bioenhancement would have even more relevance for this category of intervention. Rather than increasing the levels of a disposition that an individual perceives himself to possess in insufficient levels, where this intensifies a desire to act upon this disposition, the targeting of counter-moral emotions would seek to mitigate such emotions. This category would therefore include any morally relevant emotion that an individual perceives to be the cause of her desire to act in a way that is odds with a desire that she would prefer to have, and to act upon. Douglas suggests moral bioenhancement as a means of reducing the tendency for aggression as well as the mitigation of racial bias. There are, however, many possibilities that could be explored and their inclusion as possible contenders for moral bioenhancement would, of course, depend upon whether these dispositions are susceptible to safe biological alteration.

We may then construct a check list of requirements by utilising insights from Ekstrom's theory to assess a particular desire as a potential contender for some form of moral bioenhancement intervention[95]. I would argue that any moral bioenhancement intervention that meets the following requirements could be regarded as an enhancement of self-determination, rather than a threat to the latter. To formulate this slightly differently, in the case of an individual considering undergoing a moral bioenhancement intervention, to avoid impacting on the individual's autonomy, an intervention would have to avoid the outcomes listed below.

1) The intervention must not result in any changes to any of the authorised preferences that form part of an individual's cohering network, where this network is the means by which he has identified a particular desire as problematic to him. In other words, an intervention must not produce any changes to the integral self or the core identity of the individual as this is the source of the self in self-determination.

[95] The safety and efficacy of the intervention would have to be established, of course. This would be a requirement of any intervention that acts upon human physiological processes. Establishing efficacy would include being able to ensure that interventions are able to target specific dispositions, thus producing localised rather than global changes.

2) In connection with the above requirement, the intervention must not change any of the components that characterise an individual's conception of what is true and good, as this too forms part of her integral self and is the basis upon which she has identified a particular desire as problematic to her.
3) An intervention may not serve to bring about congruence between a desire and a preference where doing so would more easily support actions that would impact upon the negative liberty and autonomy of others. This is a protective addition informed by moral libertarian ideals[96], that would be a necessary stipulation for relatively unusual cases in which an individual has a desire that is at odds with a network of cohering preferences that could be identified as pathological. For example, an individual may have a desire to commit certain acts such as killing, maiming or raping another, and this desire may be informed by his cohering network of preferences. However, at times, he may experience an anomalous preference that he not have such desires to act in this way. This could then produce unwanted feelings of guilt and inner conflict that could be eradicated by an intervention that would increase his desire to act in accordance with his network of cohering preferences, with the result that the individual acts without further reservation in terms of his desire to kill, maim or rape. Thus, a certain normative component would be required in cases of moral bioenhancement interventions that aim to bring about congruence between desires and preferences. This would be premised upon the normative, but uncontroversial, assumption that certain actions committed upon unwilling subjects are undesirable, and therefore, the possibility of their occurring should be minimised. This component would, however, be a minimal one in order to avoid charges of paternalism and uphold respect for self-determination. However, if we operate with an interpretation of moral bioenhancement akin to the formulation provided at the end of chapter 2, then such a possibility would be excluded by definition.
4) An intervention may not change a desire that is supported by a preference that is enduring, defensible, or, that the individual is comfortable with. This is because any desire that meets these characteristics would, in-all-likelihood, be a component of the individual's integral self. While it seems highly unlikely that an individual would seek to change a desire that meets these three requirements, there could be cases in which an individual lacks adequate self-insight or is in denial about their integral self. This requirement would therefore

[96] By moral libertarianism, I propose a particular interpretation of libertarianism, referring to the acts that a thinker such as Mill would argue justify state intervention and thus impacts upon the freedom of individuals. In other words, acts that impact upon the liberty of others due to the fact that they are characterised by force, threat, manipulation or coercion. Thus, any intervention that would strengthen a desire to act in a way that that is forceful, threatening, manipulative, or coercive would be excluded as an option from which to choose.

presuppose the presence of a minimum level of rationality, self-insight and honesty on the part of the individual, qualities generally associated with the capability or conditions for being able to exercise autonomy in the first place. Furthermore, this requirement could only be fulfilled through the self-reporting of the individual and not a third party, as this would introduce the possibility of abuse and exploitation akin to the dangers associated with positive accounts of liberty discussed in section 5.2 of the first part of this chapter.

5) On the grounds of producing an increase in her autonomy, an intervention may only be utilised to alter a desire that has produced ongoing distress for an individual due to its persistence, its conflict with her cohering network of preferences, and its strength in causing her to frequently act upon it. Regarding the latter, this would refer to cases in which the individual regularly acts upon an unauthorised preference when her preference is rather for an opposing desire to be the basis upon which she acts. This is the most stringent requirement due to the fact that a desire may only be considered to be a suitable contender for moral bioenhancement, *on the grounds of increasing her autonomy*, if it is one that regularly compels her to act.

It must be remembered that what I am attempting to provide with the above list, is a set of requirements or conditions for an intervention to be considered as an enhancement of autonomy. An individual could, of course, voluntarily choose to undergo moral bioenhancement – if it were safe, effective, and available as an option in a future where the ethical concerns had been addressed – and not meet all of these requirements. However, I would argue that without meeting the above requirements there would be cause for concern regarding the authenticity of her decisions after such an intervention, and therefore, possible consequences for her autonomy.

I posit that, despite all of the above considerations, very few individuals would avail themselves of moral bioenhancement interventions as a means of addressing perceived moral deficits. It would be far more likely that such interventions would be most relevant for those with pathological behaviours and addictions that have moral consequences for the individual, where such behaviour is the product of a desire that an individual would prefer to not act upon, or for the treatment of cases of moral psychopathy. However, one of the aims of this analysis has been to investigate whether moral bioenhancement would be a violation of autonomy *in toto*. My finding is that whilst certain interventions would result in obvious threats to autonomy, a specific type of case would produce the opposite outcome and would, in fact, increase the autonomy of the individual, where

the latter is associated with the individual's ability to determine themselves in as authentic manner as possible.

5.13 Concluding remarks

My main aim with this analysis has been to show that moral bioenhancement would not, necessarily, produce global impacts upon human moral autonomy. In other words, contrary to how it is sometimes presented in the literature, what is at stake is neither an absolute loss of autonomy, nor a matter of no loss of autonomy, whatsoever. The matter is more complex than this binary would suggest. Rather, the impact upon autonomy, and thus morality, would depend on a variety of factors. It would primarily be informed by the specific nature of the intervention, where this refers to what is being targeted and how it is being targeted. These matters are entirely empirical in nature and their explication – were moral bioenhancement to ever become a possibility in the way that it has been envisaged in the literature – would depend upon the relevant scientific expertise.

Thus, while the term moral bioenhancement, in the sense that I defined it at the end of chapter 2, is used to refer to any intervention that would result in discernible improvements in moral conduct, if this were ever to become a possibility there would presumably be multiple mechanisms by which this would be achieved, ranging from less to more biologically invasive. The term, therefore, encompasses a host of interventions whose effects on autonomy would represent a continuum ranging from less problematic to more problematic. However, the term is used as if it were referring to singular interventions, rather than collections of highly diverse potential interventions, in order to facilitate coherent discussion and be able to engage with the philosophical ramifications. Thus, the moral bioenhancement debate is characterised by an ephemeral quality: it is a philosophical inquiry regarding the ethical status of a practical endeavour that may or may not ever be scientifically possible and must, therefore, be discussed in somewhat over-simplified terms. Furthermore, because of this speculative nature of the debate and the interventions that have been proposed, one is required to sidestep the scientific or practical parts of the 'problem' and investigate the ethical status of the concern for autonomy in terms of possible outcomes, as I have done, whilst still attempting to do so in a coherent manner. For this reason, I have grouped interventions in terms of the potential outcomes or impacts upon autonomy that they could be expected to produce, rather than focussing upon interventions grouped by type.

In the introduction to this second half of chapter 5, I presented a list of those outcomes that have been identified in the literature as problematic for autonomy. What I aimed to do in the rest of this chapter was to provide more substantial arguments as to why these outcomes would be problematic. Specific freedom-related arguments may be made for why the first and last items on the list – compulsory or covertly administered interventions and those interventions that would produce irreversible effects – would be morally problematic. However, in terms of the middle four outcomes, these are concerns that must be explicated by means of autonomy arguments, and thus, autonomy theories. I have, therefore, attempted to give more substance to these four potential outcomes by means of Ekstrom's theory of autonomy.

The concern that an intervention could excessively heighten an emotional response, thus producing compulsive behaviour, where the latter may be understood as acting without any cognitive reflection, is akin to the concern that an intervention could produce unsupported preferences, as discussed in section 5.11.2. In other words, this concern is directed at any intervention that creates a new desire which then becomes a preference that is sufficiently compelling to cause action, despite not having been authorised by the integral self in the stipulated manner. The third, fourth and fifth concerns listed in the introduction are different formulations of the same concern that is investigated in section 5.11.3, namely, the concern that interventions could change the integral self or psycho-moral identity of the individual. An intervention that substantially altered the identity, core beliefs or value system of an individual would, in turn, impact the means by which the individual is able to assess and accept any of these changes, with hidden changes doing so in a more decisive manner. When explicated in Ekstrom's terminology, this refers to any intervention that substantially alters the cohering network of preferences that has been critically evaluated against the individual's conception of what is true and good, because this network is the means by which she either authorises or rejects her desires and preferences, and is therefore the condition for her ability to exercise self-determination. Interventions that produced the kinds of outcomes described in sections 5.11.2 and 5.11.3 would be particularly problematic where there is pre-existing congruence between the desires and preferences of the individual, and thus, a lack of inner conflict experienced by him, because this would indicate that he has been acting in a way that is self-determined. Thus, such an intervention would thwart his self-determination in some way.

On the other hand, situations in which an individual's desires and preferences are not in congruence, and, more specifically, when she finds herself acting regularly upon unauthorised preferences, are an entirely different matter. An example of such a situation would be an individual

who feels gripped by compulsive desires that are alienating and distressing to her and that produce negative effects when she acts upon them, thus interfering with both who she takes herself to be and who she aspires to be. In such cases, if the individual then freely chose to undergo an intervention that would ameliorate such desires, to the level that she no longer felt motivated to act upon them, but was rather able to act upon those preferences that were authorised by her, I would argue that this would be a positive outcome for her and would represent an increase in the level of autonomy that she experiences. I would also argue that the entirely different issue of whether or not her post-enhanced actions would possess equal moral worth in comparison to their pre-morally bioenhanced state, would not be as urgent for her as her lack of self-determination.

Chapter 6 – Conclusion

6.1 Introduction and overview

In this dissertation, my aim has been to provide a comprehensive philosophical and ethical investigation of moral bioenhancement with a particular focus on the concern for moral autonomy that is allegedly posed by moral bioenhancement. My research has indicated that a coherent explication of the above aims and concern is both an empirical and a speculative matter. In other words, the nature of the phenomenon of moral bioenhancement is such that we must extrapolate from existing knowledge to make predictions regarding the safety and efficacy of moral bioenhancement, its social and political consequences, and its implications for both personal autonomy and universal moral autonomy. In addition, making sense of moral bioenhancement from a philosophical and ethical perspective requires conceptual clarification and analysis, an explication of the implicit meta-ethical assumptions and the challenging task of trying to pinpoint both the nature of morality itself and the conditions for its exercise.

At heart, arguments in favour of moral bioenhancement and opposing arguments that focus on the concern for moral autonomy represent a fundamental clash in value systems that is irresolvable. If the concern regarding the threat of ultimate harm raised by Persson and Savulescu is valid, and their attribution of this threat to problematic human moral dispositions and behaviour is correct, then, as mentioned in the previous chapter, the fundamental disagreement pertains to what one would be willing to sacrifice to ensure, or at least increase, our chances of survival. The fact that Harris has clearly argued that he "like so many others would not wish to sacrifice freedom for survival" (2016:74-75), indicates both where he stands on the matter as well as the irresolvable nature of this disagreement. His statement here clearly indicates that he regards freedom[97], in terms of the role that it plays in morality, as more important than survival which would indicate that freedom/autonomy comes close to enjoying the status of being absolutely and intrinsically valuable in his eyes. Such a perspective is difficult to substantiate with argumentation as it lies at the foundational level of an individual's belief or value system. In other words, we cannot delve deeper or below such a belief to justify it, one either subscribes to such a belief or one doesn't, which is why I take the moral bioenhancement debate, as it is currently approached and framed, to be irresolvable.

[97] Used in the sense of internal freedom, positive liberty or autonomy, as explicated in chapter 5.

For this reason, I have not attempted to develop Persson and Savulescu's explicitly consequentialist argument in support of moral bioenhancement as such an argument would hold no sway with those whose arguments and responses in the literature aim to protect 'goods' that are non-consequentialist in nature. While I am not positing that thinkers such as Harris, and those who share his concern for moral autonomy, disregard the importance of consequences in assessing the ethical status of proposed interventions, Harris' statement above clearly indicates that he attributes more value to the non-consequentialist, or intrinsic, 'good' of freedom, where this forms the conditions for morality, than to the consequentialist or instrumental good of survival. This is why I have chosen to engage with the concern for moral autonomy on its own terms, and from within, rather than simply responding with a consequentialist counter-argument which would be summarily rejected on the basis of the above-mentioned fundamentally held belief regarding the importance of freedom and the role it plays in morality. Responding in this way required an in-depth discussion and explication of the nature of autonomy itself, in order to ascertain whether the concern for moral autonomy posed by moral bioenhancement, is, in fact, a legitimate one.

In this brief conclusion to my dissertation I will discuss, in section 6.2, how my research has contributed to the moral bioenhancement debate. In section 6.3 I will then discuss what I posit will be the most likely application of moral bioenhancement as well as my suggestions for future research in this field. I will then conclude in section 6.4 with a final comment on the relationship between morality and autonomy.

6.2 Contributions and findings

My dissertation is the product of an extensive exploration of the literature which included all articles, chapters and books that have been published on the subject since 2008, up until the current time of writing. From this exploration, I synthesised the insights of these publications into a discussion of the subject that may now serve as a comprehensive ethical and philosophical overview that is contained within once source. This will serve as a resource to researchers who wish to investigate the area without having to expend the time required to peruse the vast and disparate literature on the subject.

In my synthesis of the literature I have aimed to bring clarity to the debate in several ways. Firstly, in the introduction to this dissertation I delineated various ways in which an investigation of moral bioenhancement could be approached. While I have approached the problem from a particular

perspective, this delineation serves to indicate alternative approaches that could be taken by other researchers interested in the field.

Secondly, I have tried to bring further clarity to the debate in terms of how I have organised the different problems that would have to be resolved for moral bioenhancement to become a coherent possibility. In this regard, my discussion in chapter 2 identified the most salient conceptual and practical problems associated with moral bioenhancement. Both definitional problems and disagreements regarding the content of morality were shown to be conceptual in nature while addressing the question of whether or not moral bioenhancement could ever be scientifically feasible requires empirical research. However, for this to commence in a coherent manner, the above-mentioned conceptual problems must be resolved as we would have to agree on what it is that should be targeted, and, we would have to be correct in our identification of these targets for moral bioenhancement to actually achieve its aims. Thus, the problem of the science of moral bioenhancement is also conceptual in nature. I also aimed to further the debate by providing a definition at the end of chapter 2 that was sufficiently rigorous and comprehensive to encompass a variety of concerns.

Thirdly, I have attempted to bring further clarity to the debate through the way in which I have grouped the various ethical concerns that have been lodged in the literature. Although categorising the ethical concerns into those that are practical or consequentialist in nature and those that are non-consequentialist belies the overlap between the two, it is nevertheless a coherent and helpful distinction as it clearly illustrates the varying levels of threat and risk that moral bioenhancement poses. In this regard, my research indicates that practically, there are risks at both individual and societal levels, while the concern for moral autonomy is a concern for something that is decidedly abstract and ephemeral in nature. What is captured by this latter concern is a related set of phenomena that I refer to as moral autonomy, where this term refers to one's sense of self as a moral being, one's personal moral identity or moral authenticity and one's perception of universal human morality as such.

In terms of this latter concern, my aim was to use the larger scope of a dissertation to explore this area in greater depth than the somewhat cursory way in which it has been discussed in the literature. In chapter 4b I provided a detailed discussion of the concern for moral autonomy, as voiced in the literature, while both parts of chapter 5 represent my unique contribution to explicating and providing depth to this concern. Ekstrom's coherence theory of autonomy has great potential for

use in areas of bioethics that require a richer account of the notion, despite the fact that, to my knowledge, her work has not yet been used in this field. I will discuss this matter further in section 6.3. In particular, Ekstrom's theory provides an interpretation of autonomy that predicates the notion on an authenticity that is justified by a structural relationship between preferences and attitudinal states that are internal to the self, which is particularly relevant for the issue at hand. In addition, I was able to utilise her account to illustrate the connection between threats to personal identity and moral autonomy. Through my application of the coherence theory of autonomy in combination with insights from hierarchical accounts, I came to the conclusion that the threat posed to autonomy by moral bioenhancement is not absolute, despite being presented this way in the literature; rather, it depends on a number of factors of which the nature of the intervention and the interpretation of autonomy are the most salient. My argument in this regard was that while a number of outcomes of moral bioenhancement interventions could pose a distinct threat to moral autonomy, certain outcomes could actually achieve the opposite, resulting in an increase in the level of autonomy experienced by individuals.

6.3 Probable applications and suggestions for future research

In both the literature and my dissertation, the discussion has addressed the possibility of the *enhancement* or improvement of psycho-moral dispositions, rather than the *treatment* of pathological functioning. In other words, the focus has been on the ethical problems associated with heightening dispositions such as empathy or altruism and a sense of justice that exist within a range of normalcy, where normalcy would be associated with those levels that fall within a particular distribution found in the population. Of course, the division between what is deemed to be a normal level of empathy, for example, and a low or pathological level of empathy, is not self-evident or given in some way. This is one of the problems with the issue regarding when treatment becomes enhancement. However, the nature of Persson and Savulescu's argument is such that they do not need to engage with the coherence of the treatment/enhancement distinction. This is because they do not argue that we must only seek to treat the psycho-moral dispositions of those with low or pathological levels; rather, they have suggested that everyone should be subject to an improvement of their levels of these dispositions in order to raise them to the levels of those members of society that are regarded as displaying the highest levels thereof.

However, if we agree with the prevailing view in the literature that the kind of compulsory or universal moral bioenhancement interventions suggested by Persson and Savulescu are ethically problematic then we can enquire as to what the real-world prospects of the moral bioenhancement

project would be. It seems that even if it became possible to safely and effectively elevate the levels of the above-mentioned psycho-moral dispositions or to mitigate counter-moral emotions, as discussed by Douglas, the application of such interventions would, in all likelihood, be limited to certain cases. In particular, it is my contention that such interventions would be limited to two kinds of possible contexts.

Firstly, individuals who have problematised their lower levels of such dispositions or high levels of counter-moral emotions, due to the fact that their existing levels have negatively impacted their lives or dealings with others producing negative overall consequences, may seek to avail themselves of such interventions, if it has been established that improving these levels will bring about general improvements in their lives. In other words, in the same way that individuals freely choose to avail themselves of pharmacological medications for conditions such as depression and anxiety, to improve the quality of their lives, they may also choose to heighten their levels, of empathy, for example, if they believe that doing so would achieve the same end. If the safety and efficacy of such interventions had been established, it is possible that the number of individuals opting to undergo such interventions may be substantial. Furthermore, in cases where individuals had problematised their existing levels of these dispositions to the extent that they freely chose to undergo a moral bioenhancement intervention as a means of improving their lives, it is likely that doing so would achieve congruence between their preferences and integral self as discussed in section 5.12 of chapter 5b. Therefore, in such cases, undergoing such an intervention would possibly increase their autonomy as self-determination for the reasons presented in section 5.12 of chapter 5b.

I would argue that the second most likely application of moral bioenhancement would be to make such interventions available to those individuals who possess pathological or low levels of such dispositions where this has led them to engage in criminal behaviour for which they have been incarcerated. This matter was briefly discussed in section 3.3.1 of chapter 3. The volume of discussion in the literature regarding this possible application of moral bioenhancement has grown considerably in recent years[98]. While the potential for abuse of such interventions is vast, what has actually been suggested, in this regard, is less controversial than one would suppose upon first consideration. The most prevalent possibility suggested in the literature would be to offer criminals a choice between incarceration or a moral bioenhancement intervention if their behaviour could be

[98] References here are too numerous to list due to the rapid expansion of this research area; however, for some examples see Shook, 2012; Selgelid, 2014; Douglas, 2014; Curtis, 2012; Horstkötter et al., 2012; Wiseman, 2014; Beck, 2015, Caouette, 2015; Barn, 2016.

attributed to a deficiency for which there was an available treatment. As pointed out by Douglas, we already do this in certain cases when we offer chemical castration to individuals found guilty of paedophilia in exchange for early release or parole (2014:103). One of the objections to this possible application, however, has been the argument that a choice between incarceration and a biomedical intervention is not a genuine choice. As mentioned in section 3.3.1 of chapter 3, in terms of Selgelid's continuum of freedom that includes degrees of encouragement and discouragement for undergoing moral bioenhancement: the more compelling the discouragement of failing to partake in moral bioenhancement, and thus, the more individuals feel they will be disadvantaged by punitive measures, the more their freedom will be compromised (2014:216). I would argue, however, that this second application of moral bioenhancement would be the most probable application of the types of interventions that have been discussed in this dissertation and that it therefore warrants further ethical investigation and empirical research.

An entirely different suggestion for further research of other themes discussed in this dissertation would be to develop the checklist of intervention outcomes associated with the protection of autonomy presented in section 5.12 of chapter 5b into a more practical set of guidelines that may be utilised in bioethics contexts that require a richer account of autonomy. In bioethics, autonomy is generally important in terms of its role in securing patient rights, particularly their right to informed consent. In terms of its role in such contexts, a richer account of autonomy is not relevant as the concept serves as tool of protection or empowerment for patients, rather than as some abstract quality or capability of individuals that requires protection itself. In terms of this distinction, a richer account of autonomy would be of use predominantly in the case of any interventions that threaten this quality or capability of individuals to be autonomous on a deeper or global level, or where patients are incapacitated and unable to make decisions pertaining to their health. Regarding the former possible application, this would most likely be applicable in cases of more invasive neurobiological interventions where there is concern that the intervention may produce unwanted impacts and changes to the identity of individuals.

6.4 A final comment on the relationship between morality and autonomy

The concern for moral autonomy in the literature is a difficult one to make sense of as it fluctuates between two formulations that are seemingly distinct. On the one hand, it is sometimes framed as the argument that moral bioenhancement, by making it more likely that we would act in a particular manner – more moral, in this case – would lessen our autonomy in a general sense. In other words, post-moral bioenhancement, we would be less autonomous than we were before the intervention;

our behaviour would have become predetermined in some way. This would be the concern that moral bioenhancement could amount to a form of behaviour control, or, even in cases of voluntary moral bioenhancement, that it will be a form of moral compulsion. On the other hand, the concern is also framed as an argument that moral bioenhancement would erode morality as such. The underlying argument here, is that if authentic morality requires that our behaviour be unpredictable, so that it is unclear what moral choices we will make in any given situation, an intervention that reduces this unpredictability, will also reduce or erode our morality.

When formulated in this way, it is clear that the two formulations are related, and that the second formulation follows on from the first one, or that the first formulation is a justification of the second. They are, nevertheless, concerns pertaining to two different phenomena. The first formulation of the concern could be levelled at any intervention that would make it more likely that an individual acts in a particular manner, or is directed towards a particular behaviour or choice. Thus, the first formulation could be an argument that may also be levelled at interventions with non-moral outcomes; it is not specific to moral bioenhancement. In other words, if the concern is that behaviour becomes more determined in some way after an intervention, and, in keeping with the nature and aim of a particular intervention, the concern for autonomy could be levelled at interventions that increase an individual's sense of cooperation or compliance in general, or even those interventions that introduce a new or improved ability in individuals, thus making it more likely that they will direct their lives towards utilising this ability. In terms of the latter, having a particular talent or ability may produce a sense of determination in some way. If I 'find' myself to be a talented singer I am more likely to feel 'compelled' to make something of this talent, such as attempting to have a career in music despite the sacrifices this may entail. If I possess mathematic acumen, I will be more likely to choose certain careers, and thus, an important portion of my life will have been determined in some sense by my abilities.

However, despite this, we never encounter such autonomy concerns in the literature. In other words, while there are a variety of ethical concerns that have been raised towards, for example, cognitive bioenhancement or the bioenhancement of personal abilities such as musicianship or athletic ability, these bioenhancements do not elicit concern for the autonomy of the individuals if they freely choose to undergo such interventions. Therefore, the prevalence of the concern for autonomy in the moral bioenhancement literature and its absolute lack of presence in the bioenhancement literature in general indicates that we are, indeed, dealing with something entirely different in kind.

In my dissertation, I have chosen to focus predominantly on the first formulation of the concern by arguing that thinkers such as Harris overestimate the extent to which certain individuals possess autonomy to begin with. Simply put, I have argued that where there is a lack of congruence between how an individual acts and how she would wish to act, then she is in a state of compromised autonomy to begin with. Formulated differently, the extent to which an individual has problematised his morally relevant behaviour is an indication that an intervention that mitigates this should not be regarded as eroding his autonomy; it is rather, an affirmation of the latter. If one accepts these arguments then there is not necessarily a need to engage with the second formulation of the concern. However, the matter is complicated by the nature of the second formulation and the fact that, as mentioned above, in the case of moral bioenhancement we are dealing with something different in kind. For this reason, the second formulation is more challenging and possibly irresolvable, as, if one delves down to its foundations, it becomes evident, once more, that it represents a fundamental value clash.

Harris' argument that moral bioenhancement would eradicate our 'freedom to fall' which is a precondition for morality, is clearly a version of the second formulation of the concern for moral autonomy. As mentioned above, in my dissertation I have largely chosen to engage with the first formulation of the concern as it is the justification or argument that supports the second formulation. However, turning briefly now to the second formulation, if we explore what lies at the heart of Harris' argument, it is his concern for morality as such. In fact, as mentioned in section 6.1, for Harris, the value of morality is absolute. His statement regarding the fact he wouldn't "wish to sacrifice freedom for survival" (Harris, 2016:74-75) should rather be reformulated as not wishing to sacrifice morality or moral freedom for survival. The fact that he persists with this proposition, arguing that it would stand even in cases where individuals had freely chosen to undergo moral bioenhancement, indicates that this is his deepest concern.

The matter of whether or not moral bioenhancements would decidedly alter human morality as such, in the way that thinkers such as Harris fear, is not possible to definitively ascertain. However, for the reasons discussed at various points in this dissertation, it is my contention that the concerns in this regard are exaggerated. However, to take an entirely different approach to this matter, and in conclusion, I would argue that if one engages directly with the second formulation of the concern, then both the relevance and strength of Harris' argument – or the concern for morality in general – is dependent upon how we ascribe value to the phenomenon of morality. If we view morality only,

or primarily, in terms of its intrinsic value, as Harris' statement above clearly indicates he does, then it is something to be protected at all costs, as if it were an item of such value that we should seek to preserve it even at the cost of our survival as a species. In other words, it would be better not to survive if we do so at the expense of our morality. However, if one holds the view that a considerable part of why we value morality lies in its instrumental value – it enables harmonious coexistence and societal functioning, it is a protective mechanism against potential harms that others could inflict on us, and a world with the presence of morality will be more pleasant, and thus preferable to a world without it – then one will rather support the view that morality should serve the ends of human beings, rather than human beings serving the ends of morality. In other words, the argument here would be that if we can improve upon morality then it will better serve the instrumental ends required of it. Why should morality, therefore, be reified or treated as something that should remain unchanged? This argument, however, is indicative of a consequentialist perspective. Thus, we are once again faced with the impasse between consequentialism and non-consequentialism. This fundamental disagreement aside, the problem with framing the debate in terms of the protection of morality, as such, is that in doing so, one elevates an abstract phenomenon, such as morality, to a status that, I would argue, occurs at the expense of the concrete individual whose freedom to choose not to fall so often, or to be a more moral person, is disregarded.

Bibliography

Agar, N. 2010. Enhancing Genetic Virtue?. *Politics and the Life Science*, 29(1):73-75.

Agar, N. 2014. A question about defining moral bioenhancement. *Journal of Medical Ethics*, 40(6):369-370.

Agar, N. 2015a. Moral bioenhancement is dangerous. *Journal of Medical Ethics*, 41(4):343-345.

Agar, N. 2015b. Moral Bioenhancement and the Utilitarian Catastrophe. *Cambridge Quarterly of Healthcare Ethics*, 24(1):37-47.

Ainslie, G. 2001. *Breakdown of will*. Cambridge: Cambridge University Press.

Albert, H. 1968. *Traktat über kritische Vernunft*. Heidelberg: Mohr (Siebeck)

Andreadis, 2010. The tempting illusion of genetic virtue. *Politics and the Life Sciences*, 29(1):76-78.

Aristotle. 2004. *The Nichomachean Ethics*. J.A.K. Thomson (tr.). London: Penguin Books.

Arnhart, L. 2010. Can virtue be genetically engineered? *Politics and the Life Sciences*, 29(1):79-81.

Baertschi, B. 2014. Neuromodulation in the service of moral enhancement. *Brain Topography*, 27(1):63-71.

Barilan, Y.M. 2015. Moral Enhancement, Gnosticism, and Some Philosophical Paradoxes. *Cambridge Quarterly of Healthcare Ethics*, 24(1):75-85.

Barn, G. 2016. Can medical Interventions Serve as 'Criminal Rehabilitation'? Neuroethics. [Online.]. Available: https://link.springer.com/content/pdf/10.1007%2Fs12152-016-9264-9.pdf. [2017 August 17].

Baron-Cohen, S. 2003. *The Essential Difference: Male and Female Brains and the Truth about Autism*. New York: Basic Books.

Baylis, F. 2013. I am who I am: On the perceived threats to personal identity from deep brain stimulation. *Neuroethics*, 6(3):513-526.

Beauchamp, T.L. 2015. Are we unfit for the future? *Journal of Medical Ethics*, 41(4):346-348.

Beck, B. 2015. Conceptual and practical problems of moral enhancement. *Bioethics*, 29(4):233-240.

Benjamin, J. et al. 1996. Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nature Genetics*, 12:81-84.

Berlin, I. 1969. *Four Essays on Liberty*. London: Oxford University Press.

Blackford, R. 2010. Genetically engineered people. *Political Life Science*: 29(1):82-84.

Bostrom, N. & Roache, R. 2008. Ethical Issues in Human Enhancement, J. Ryberg, T. Petersen & C. Wolf (eds.). *New Waves in Applied Ethics*. Pelgrave Macmillan:120-152.

Bostrom, N. 2003. *The Transhumanist FAQ*. Version 2.1. [Online.]. Available: http://www.transhumanism.org/resources/FAQv21.pdf [2013, October 9].

Bronstein, J. 2010. Objecting to the Genetic Virtue Program. *Politics and the Life Scie*nces, 29(1):85-87.

Brooks, T. 2012. Moral Frankensteins. *American Journal of Bioethics: Neuroscience*, 3(4):28-30.

Bruni, T. 2011. The Ambivalence of Moral Psychology. *American Journal of Bioethics: Neuroscience*, 2(4):13-15.

Brunner, H.G., Nelen, M.R., Breakefield, X.O. et al. 1993. Abnormal behaviour associated with a point mutation in the structural gene for Monoamine Oxidase A. *Science*, 262(5133):578–580.

Bublitz, C. 2016. Moral Enhancement and Mental Freedom. *Journal of Applied Philosophy*, 33(1):88-106.

Buchanan, A., Brock, D.W. Brock, Daniels, N. & Wikler, D. (eds.). 2009. *From Chance to Choice Genetics and Justice*. New York: Cambridge University Press.

Buchanan, A. 2011. *Beyond Humanity*. Oxford: Oxford University Press.

Cadoret, R.J. 1978. Psychopathology in adopted-away offspring of biologic parents with antisocial behaviour. *Archives of General Psychiatry*, 35:176–184.

Cambier, H. 2006. Is Popper's Philosophy Anti-Foundationalist?, I. Jarvie, K. Milford, D. Miller (eds.). *Karl Popper: A Centenary Assessment Volume II Metaphysics and Epistemology*. Aldershot: Ashgate Publishing Limited:145-156.

Caouette, J. 2015. *On the Moral Permissibility of Passive Moral Enhancement: Comment on "Do Means Matter Morally?*". [Online]. Available: https:// www.academia.edu/15195473/ On_the_Moral_Permissibility_of_Passive_Moral_Enhancement [2016, March 27].

Carter, J.A. & Gordon, E.C. 2013. On cognitive and moral enhancement: a reply to Savulescu and Persson. *Bioethics*, 29(3):153-161.

Casal, P. 2015. On not taking men as they are: reflections on moral bioenhancement. *Journal of Medical Ethics*, 41(4):340-342.

Caspi, A. & McClay, J. 2001. Evidence that the cycle of violence in maltreated children depends on genotype. *Science*, 297: 851–854.

Chan, S. & Harris, J. 2011. Moral enhancement and pro-social behaviour. *Journal of Medical Ethics*, 37(3):130-131.

Christen, M. & Narvaez, D. 2012. Moral Development in Early Childhood is Key for Moral Enhancement. *American Journal of Bioethics: Neuroscience*, 3(4):25-26.

Christman, J. 1989. *The Inner Citadel: Essays on Individual Autonomy*. New York: Oxford University Press.

Christman, J. & Anderson, J. 2005. Introduction, in J. Christman and J. Andersen (eds.). *Autonomy and the Challenges to Liberalism*: New Essays. New York: Cambridge University Press:1-23

Clausen, J. 2010. Ethical brain stimulation – neuroethics of deep brain stimulation in research and clinical practice. *European Journal of Neuroscience*, 32(7):1152-1162.

Crockett, M., et al. 2008. Serotonin modulates behavioural reactions to unfairness. *Science*, 320(5884):1739.

Crockett, M., Clark, L., Hauser, M.D., et al. 2010a. Serotonin selectively influences moral judgement and behaviour through effects on harm aversion. *Proceedings of the National Academy of Science*,107:17433–8.

Crockett, M., Clark, L., Hauser, M.D., & Robbins, T.W. 2010b. Reply to Harris and Chan: Moral judgment is more than rational deliberation. *Proceedings of the National Academy of Science*, 107(5):E184.

Crockett, M., et al. 2010c. Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion. *Emotion*, 10(6):855–62.

Crockett, M.J. 2014. Moral bioenhancement: a neuroscientific perspective. *Journal of Medical Ethics*, 40(6):370-371.

Crowe, R.R. 1974. An adoption study of antisocial personality. *Archives of General Psychiatry*, 31:785–791;

Crutchfield, P. 2016. The Epistemology of Moral Bioenhancement. *Bioethics*, 30(6):389-396.

Cunningham, W.A., Johnson, M.K., Raye, C.L. et al. 2004. Separable neural components in the processing of black and white faces. *Psychological Science*, 15:806–813.

Curtis, B.L. 2012: Moral enhancement as rehabilitation? *American Journal of Bioethics Neuroscience*, 3(4):23-24.

Daniels, N. 2000. Normal Functioning and the Treatment-Enhancement Distinction. *Cambridge Quarterly of Healthcare Ethics*, 9(3):309-322.

Davidson, D. 1980. How is Weakness of the Will Possible? *Essays on Actions and Events*. Oxford: Oxford University Press:21-41.

De Almeida, R.M.M., Ferari, P.F., Parmigiani, S. et al. 2005. Escalated aggressive behaviour: Dopamine, serotonin and GABA. *European Journal of Pharmacology*, 526:51–64.

De Dreu, C. et al. 2011. Oxytocin Promotes Human Ethnocentrism. *Proceedings of the National Academy of Science*, 108:1262-6.

Deep Brain Stimulation for Movement Disorders. 2016. [Online]. Available: http://www.neurosurgery.pitt.edu/centers-excellence/epilepsy-and-movement-disorders-program/deep-brain-stimulation-movement-disorders [2016, October 26].

DeGrazia, D. 2005. Enhancement technologies and human identity. *Journal of Medical Philosophy*, 30(3):261-283.

DeGrazia, D. 2014. Moral enhancement, freedom, and what we (should) value in moral behaviour. *Journal of Medical Ethics*, 40(6):361-368.

De Melo-Martin, I. & Salles, A. 2015. Moral bioenhancement: Much Ado About Nothing? *Bioethics*, 29(4):223-232.

De Waal, F. 1996. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press.

De Waal, F. 2006. *Primates and Philosophers: How Morality Evolved*. In S. Macedo & J. Ober (eds.), Princeton, NJ: Princeton University Press.

Douglas, T. 2008. Moral Enhancement. *Journal of Applied Philosophy*, 25(3):228-245.

Douglas, T. 2013. Moral enhancement via direct emotion modulation: a reply to John Harris. *Bioethics*, 27(3):160-168.

Douglas, T. 2014. Criminal Rehabilitation Through Medical Intervention: Moral Liability and the Right to Bodily Integrity. *Journal of Ethics*, 18:101-122.

Drake, N. 2016. Is moral bioenhancement dangerous?. *Journal of Medical Ethics*, 42(1):3-6.

Dworkin, G. 1970. Acting Freeling. *Nous*, 4:367-383.

Dworkin, G. 1976. Autonomy and Behaviour Control. *Hastings Center Report*, 6(1):23-28.

Dworkin, G. 1988. *The theory and practice of autonomy*. Cambridge: Cambridge University Press.

Dworkin, R. 1977. *Taking Rights Seriously*. London: Duckworth.

Ebstein, R. et al., 1995. Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of novelty seeking. *Nature Genetics*, 325:783-87.

Ekstrom, L.W. 1993. A Coherence Theory of Autonomy. *Philosophy and Phenomenological Researc*h, 53(3):599-616.

Ekstrom, L.W. 1999. Review: Keystone Preferences and Autonomy. *International Phenomenological Societ*y, 59(4):1057-1063.

Ekstrom, L.W 2005a. Autonomy and Personal Integration, in J Stacey Taylor (ed.). *Personal Autonomy*. Cambridge: Cambridge University Press:143-161.

Ekstrom L.W. 2005b. Alienation, Autonomy and the Self. *Midwest Studies in Philosophy*, XXIX:45-67.

Eley, T., Lichtenstein, P. & Stevenson, J. 1999. Sex differences in the etiology of aggressive and nonaggressive antisocial behaviour: results from two twin studies. *Child Development*, 70:155–68.

Faust, H.S. 2008. Should we select for genetic moral enhancement? A thought experiment using the MoralKinder (MK+) haplotype. *Theoretical Medical Bioethics*, 29(6):397-416.

Feinberg, J. 1989. Autonomy, in J. Christman (ed.). *The Inner Citadel: Essays on Individual Autonomy*. Oxford: Oxford University Press.

Fischer, J.M. & Ravizza, M. 1998. *Responsibility and control. A theory of moral responsibility*. Cambridge: Cambridge University Press.

Focquaert, F. & Schermer, M. 2015a. Moral Enhancement: Do Means Matter Morally? *Neuroethics*, 8(2):139-151.

Focquaert, F. & Schermer, M. 2015b. *Moral Enhancement: Do Means Matter Morally? Reply to Justin Caouette*. [Online]. Available: http://philosophyofbrains.com/2015/08/25/neuroethics-symposium-on-focquaert-schermer-moral-enhancement-do-means-matter-morally.aspx. [2017, March 27].

Focquaert, F. & Schermer, M. 2015c. *Moral Enhancement: Do Means Matter Morally? Reply to Elisabeth Shaw*. [Online]. Available: http://philosophyofbrains.com/2015/08/25/neuroethics-symposium-on-focquaert-schermer-moral-enhancement-do-means-matter-morally.aspx. [2017, March 27].

Frankfurt, H.G. 1969. Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, 66(23):829-839.

Frankfurt, H.G. 1971. Freedom of the will and the concept of a person. *The Journal of Philosophy*, 58:5-20.

Frankfurt, H.G. 1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.

Frankfurt, H.G. 2002. Reply to Michael E. Bratman, in S Buss and L Overton (eds.). *Counters of Agency: Essays on Themes from Harry Frankfurt*. Cambridge, MA: MIT Press:86-90.

Franzini, A., Marras, C., Ferroli, P. et al. 2005. Stimulation of the posterior hypothalamus for medically intractable impulsive and violent behaviour. *Stereotactic and Functional Neurosurgery*, 83:63-66.

Fröding, B.E.E. 2011. Cognitive enhancement, virtue ethics and the good life. *Neuroethics*, 4(3):223-234.

Gaillot, M., Baumeister, R.F., DeWall, C.N., et al. 2007. Self-control relies on glucose as a limited energy source: willpower is more than a metaphor. *Journal of Personality and Social Psychology*, 92:325–36.

Glenn, A.L. & Raine, A. 2013. Neurocriminology: Implications for the punishment, prediction and prevention of criminal behaviour. *Nature Reviews Neuroscience*, 15(1):54-63.

Greene, J.D. et al. 2001. An fMRI Investigation of Emotional Engagement in Moral Judgement. *Science*, 293:2105:2108.

Grove, W.M., Eckert, E.D., Heston, L. et al. 1990. Heritability of substance abuse and antisocial behaviour: a study of monozygotic twins reared apart. *Biological Psychiatry*, 27:1293–1304.

Haidt, J. 2001. The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgement. *Psychological Review*, 108:814-834.

Hälbig, T.D., Tse, W., Frisina, P.G., et al. 2009. Subthalamic deep brain stimulation and impulse control in Parkinson's disease. *European Journal of Neurology* 16:493–497.

Hamilton, W. 1964. The genetical evolution of social behaviour. *Journal of Theoretical Biology*, 7:1- 52.

Harris, J. 2007. *Enhancing Evolution: The Ethical Case for Making Better People*. Princeton: Princeton University Press.

Harris, J. 2011. Moral enhancement and freedom. *Bioethics*, 25(2):102-111.

Harris, J. 2013a. Moral progress and moral enhancement. *Bioethics*, 27(5):285-290.

Harris, J. 2013b. Ethics is for bad guys! Putting the 'moral' into moral enhancement. *Bioethics*, 27(3):169-173.

Harris, J. 2013c. What it is to be good. *European Review*, 21(S1):S114-S122.

Harris, J. 2014. Taking liberties with free fall. *Journal of Medical Ethics*, 40:371-374.

Harris, J. 2016. *How to be Good*. Oxford: Oxford University Press.

Harris, J. & Savulescu, J. 2015. A Debate about Moral Enhancement. *Cambridge Quarterly of Healthcare Ethics*, 24(10):8-22.

Hart, A.J., Whalen, P.J., Shin, L.M. et al. 2000. Differential response in the human amygdala to racial outgroup vs. ingroup face stimuli. *Neuroreport: For Rapid Communication of Neuroscience Research*, 11:2351–2355.

Hauskeller, M. 2015. Being good enough to prevent the worst. *Journal of Medical Ethics*, 41(4):289-290.

Holtug, N. 1998. Creating and patenting new life forms, in H. Kuhse & P. Singer (eds.). *A Companion to Bioethics*. Oxford: Blackwell Publishers. 206-214.

Homer. 2014. *The Odyssey*. B.B. Powell (tr.). Oxford: Oxford University Press.

Horstkötter, D., Berghmans, R. & de Wert, G. 2012. Moral Enhancement for Antisocial behaviour? An uneasy relationship. *American Journal of Bioethics: Neuroscience*, 3(4):26-28.

Hubbeling, D. 2009. Pharmacology and human morality. *British Journal of Psychiatry*, 194(2):187-188.

Hughes, J.J 2013. Using neurotechnologies to develop virtues: a Buddhist approach to cognitive enhancement. *Accountability in Research: Policies and Quality Assurance*, 20(1):27-41.

Hughes, J.J. 2015. Moral Enhancement Requires Multiple Virtues. *Cambridge Quarterly of Healthcare Ethics*, 24:86-95.

Hume, D. 1978. *A Treatise of Human Nature* (2nd edition), L.A. Selby-Bigge (ed.). Oxford: Clarendon Press.

Huxley, A. 1932. *Brave New World*. London: Vintage Books.

Jebari, K. 2014. What to enhance: behaviour, emotion or disposition?. *Neuroethics*, 7(3):253-261.

Jones, D.G. 2013. The importance of realism in assessing technological possibilities: The role of Christian thinking. *Christian Perspectives on Science and Technology*, 9:1-10.

Jotterand, F. 2011. 'Virtue engineering' and moral agency: will post-humans still need the virtues?. *American Journal of Bioethics: Neuroscience*, 2(4):3-9.

Jotterand, F. 2014. Questioning the moral enhancement project. *American Journal of Bioethics*, 14(4):1-3.

Juengst, E.T. 1998. What does Enhancement Mean?, in E. Parens (ed.). *Enhancing Human Traits.* Georgetown University Press: Washington DC:29-47.

Kahane, G. & Savulescu, J. 2015. Normal Human Variation: Refocussing the Enhancement Debate. *Bioethics*, 29(2):133-143.

Kant, I. 2002. *Groundwork of the Metaphysic of Moral*. A.W. Wood (tr.). New Haven: Yale University Press.

Kass, L.R. 2002. *Life, Liberty and the Defense of Dignity: The Challenge for Bioethics*. New York: Encounter Books.

Koenigs M, et al. 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446:908–911.

Korsgaard, C. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U. & Fehr, E. 2005. Oxytocin Increases Trust in Humans. *Nature*, 435(2):673-6.

Kurzweil, R. 2001. *The Law of Accelerating Returns*. [Online]. Available: http://www.kurzweilai.net/the-law-of-accelerating-returns. [2013, October 10].

Lechner, S. 2014. Why moral bioenhancement is a bad idea and why egalitarianism would make it worse. *American Journal of Bioethics*, 14(4):31-32.

Lehrer, K. 1997. *Self-Trust: A study of Reason, Knowledge and Autonomy*. Oxford: Clarendon Press.

Locke, J. 1975. *An essay concerning Human Understanding*, P.H. Nidditch (ed.). Oxford: Oxford University Press.

Lu, C., Bharmal, A. & Suchowersky, O. 2006. Gambling and Parkinson disease. *Archives of Neurology*, 63:298.

MacIntyre, A. 2007. *After Virtue*.London: Duckworth.

Marshall, F. 2014. Would moral bioenhancement lead to an inegalitarian society? *American Journal of Bioethics,* 14(4):29-30.

Mele, A.R. 1983. "Akrasia", Reasons, and Causes. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition,* 44(3):345-368.

Menzel, E. 1974. A group of young chimpanzees in a one- acre field: Leadership and communication, in A.M. Schrier & F. Stollnitze (eds.). *Behaviour of nonhuman primates*. New York: Academic Press:83-153.

Mill, J.S. 1863. *On Liberty*. Boston: Ticknor and Fields.

Milton, 1667. *Paradise Lost*. [Online]. Available: http://triggs.djvu.org/djvu-editions.com/MILTON/LOST/Download.pdf. [2017, October 3].

Morioka, M. 2014. Some remarks on moral bioenhancement, in A. Akabayashi (ed.). *The Future of Bioethics: International Dialogues.* Oxford: Oxford University Press:120-125.

Murphy, T.F. 2015. Preventing Ultimate Harm as the Justification of Biomoral Modification. *Bioethics*, 29(5):369-377.

Neely, W. 1974. Freedom and Desire. *Philosophical Review*, 83:32-54.

Norris, C. 2010. Frankfurt on Second-Order Desires and the Concept of a Person. *Prolegomena*, 9(2):199-242.

Nozick, R. 1974. *Anarchy, State, and Utopia*. Oxford: Blackwell Publishing.

Nuffield Council on Bioethics. 2013. *Novel neurotechnologies: intervening in the brain*. London: Nuffield Council on Bioethics.

Pacholczyk, A. 2011. Moral enhancement: What is it and do we want it? *Law, Innovation & Technology*, 3(2):251-277.

Parfit, D. 1995. The unimportance of identity, in H. Harris (ed.). *Identity*. Oxford: Oxford University Press:13-45.

Persson, I. & Savulescu, J. 2008. The perils of cognitive enhancement and the urgent imperative to enhance the moral character of humanity. *Journal of Applied Philosophy*, 25(3):162-177.

Persson, I. & Savulescu, J. 2010. Moral transhumanism. *Journal of Medical Philosophy*, 35(6):656-669.

Persson, I. & Savulescu, J. 2011. The turn for ultimate harm: a reply to Fenton. *Journal of Medical Ethics*, 37(7):441-444.

Persson, I. & Savulescu, J. 2012. *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.

Persson, I. & Savulescu, J. 2013. Getting moral enhancement right: the desirability of moral bioenhancement. *Bioethics*, 27(3):124-131.

Persson, I. & Savulescu, J. 2014a. Should moral bioenhancement be compulsory? Reply to Vojin Rakić. *Journal of Medical Ethics*, 40(4):246-250.

Persson, I. & Savulescu, J. 2014b. Against fetishism about egalitarianism and in defense of cautious moral bioenhancement. *American Journal of Bioethics*, 14(4):39-42.

Persson, I. & Savulescu, J. 2015a. Reply to commentators on Unfit for the Future. *Journal of Medical Ethics*, 41(4):348-352.

Persson, I. & Savulescu, J. 2015b. The Art of Misunderstanding Moral Bioenhancement. *Cambridge Quarterly of Healthcare Ethics*, 24(1):48-57.

Persson I. & Savulescu, J. 2016. Moral Bioenhancement, Freedom and Reason. *Neuroethics*, 9(3):263-268

Pettit, P. 1997. *Republicanism: A Theory of Freedom and Government*. Oxford: Clarendon Press.

Phelps, E.A., O'Connor, K.J., Cunningham, W.A. et al. 2000. Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, 12:729–738.

Piper, A.M.S. 1985. Two conceptions of the self. *Philosophical Studies*, 48(2):173-197.

Piper, M. Undated. *Autonomy: Normative*. [Online]. Available: http://www.iep.utm.edu/aut-norm/ [2017, May 15].

Plato. 1985. *The Republic*. D Lee (tr.). Harmondsworth: Penguin Books.

Rakić, V. 2012. From cognitive to moral enhancement: a possible reconciliation of religious outlooks and the biotechnological creation of a better human. *Journal for the Study of Religions and Ideology*, 11(31):113-128).

Rakić, V. 2014. Voluntary ME and the survival-at-any-cost-bias. *Journal of Medical Ethics*, 40(4), 251-252.

Ram-Tikten, E. 2014. The possible effects of moral bioenhancement on political privileges and fair equality of opportunity. *American Journal of Bioethics*, 14(4):43-44.

Raus, K., Focquaert, F., Schermer, M., Specker, J. & Sterckx, S. 2014. On defining moral enhancement: a clarificatory taxonomy. *Neuroethics*, 7(3):263-273.

Reuter, M., Clemens, F., Walter, N.T. et al. 2011. Investigating the genetic basis of altruism: the role of the COMT Val158Met polymorphism. *Social Cognitive and Affective Neuroscience*, 6:662–8.

Riis, J., Simmons, J.P. & Goodwin, G.P. 2008. Preferences for enhancement pharmaceuticals: The reluctance to enhance fundamental traits. *Journal of Consumer Research*, 35(3):495-508.

Robichaud, P. 2014. Moral capacity enhancement does not entail moral worth enhancement. *American Journal of Bioethics*, 14(4):33-34.

Romans 7:18-20. 1980. *The Good News Bible*. Cape Town: National Book Printers.

Sandel, M.J. 2007. *The Case against Perfection*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Savulescu, J., Sandberg, A. & Kahane, G. 2011. Well-Being and Enhancement, in J. Savulescu, R. Ter Meulen & G Kahane (eds.). *Enhancing Human Capacities*. Oxford: Wiley-Blackwell:3-18.

Savulescu, J. & Persson, I. 2012. Moral enhancement, freedom and the God Machine, *Monist*, 95(3):399-421.

Savulescu, J., Douglas, T. & Persson, I. 2014. Autonomy and the ethics of biological behaviour modification, in A. Akabayashi (ed.). *The Future of Bioethics: International Dialogues*. Oxford: Oxford University Press:91-112.

Schaeffer, G.O. 2011. What is the goal of moral engineering?. *American Journal of Bioethics: Neuroscience*, 2(4):10-11.

Schechtman, M. 2010. Philosophical reflections on narrative and deep brain stimulation. *Journal of Clinical Ethics*: 21(2):133-139.

Schmidt-Salomon, M. 2007. Von der illusorischen zur realen Freiheit: Autonome Humanität jenseits von Schuld und Sühne, in K.P. Liessmann (ed.). *Philosophicum Lech. Die Freiheit des Denkens*. B. Beck (tr.). Vienna: Zsolnay:179–218.

Searle, J.R. 2001. *Rationality in Action*. Cambridge: MIT Press.

Selgelid, M.J. 2014. Freedom and moral enhancement. *Journal of Medical Ethics*, 40(4):215-216.

Shook, J.R. 2012. Neuroethics and the possible types of moral enhancement. *American Journal of Bioethics Neuroscience*, 3(4):3-14.

Simkulet, W. 2012. On moral enhancement. *American Journal of Bioethics Neuroscience*, 3(4):17-18

Singer, P. 2005. *Ethics and Intuitions. The Journal of Ethics: An International Philosophical Review*, 9(3-4):331-352.

Smeding, H.M.M., Goudriaan, A.E., Foncke, E.M.J., et al. 2007. Pathological gambling after bilateral subthalamic nucleus stimulation in Parkinson disease. *Journal of Neurology, Neurosurgery, and Psychiatry* 78:517–519.

Sober, E. & Sloan Wilson, D. 1998. *Unto Others*. Cambridge, MA: Harvard University Press.

Sorensen, K. 2014. Moral Enhancement and the Self-Subversion Objections. *Neuroethics*, 7:275-286.

Sparrow, R. 2014a. Better living through chemistry? A reply to Savulescu and Persson on 'moral enhancement'. *Journal of Applied Philosophy*, 31(1):23-32.

Sparrow, R. 2014b. Egalitarianism and moral bioenhancement. *American Journal of Bioethics*, 14(4):20-28.

Specker, J., Focquaert, F., Raus, K. Sterckx, S. & Schermer, M. 2014. The ethical desirability of moral bioenhancement: a review of reasons. *BMC Medical Ethics*: 15(67):1-17.

Spence, S.A. 2008. Can pharmacology help enhance human morality?. *British Journal of Psychiatry*, 193(3):179-180.

Stacey Taylor, J. 2005. Introduction, in J Stacy Taylor (ed.). *Personal Autonomy*. Cambridge: Cambridge University Press:1-29.

Strawson, G. 1994. The impossibility of moral responsibility. *Philosophical studies*, 75(1/2):5-24.

Terbeck, S., Kahane, G., McTavish. S. et al. 2012. Propranolol reduces implicit negative racial bias. *Psychopharmacology (Berl)*,222:419–24.

Trivers, R. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46:35-57.

Trivino, J.L.P. 2013. On the need of moral enhancement. A critical comment of "Unfit for the future" of I. Persson and J. Savulescu. *Dilemata*, 12:261-269.

Tse, W.S. & Bond, A.J. 2002. Serotonergic Intervention Affects Both Social Dominance and Affiliative Behaviour. *Psychopharmacology*, 161:324-30.

Universal Declaration of Human Rights. (1948). [Online]. Available: http://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf [2017, February 22].

Van Goozen, S.H.M & Fairchild, G. 2008. How can the study of biological processes help design new interventions for children with severe antisocial behaviour?. *Developmental Psychopathology*, 20(3):941-973.

Van Niekerk, A.A. 1980. Die grense van die kritiese rede: 'n kritiese ondersoek van die rasionaliteitsmodel van die kritiese rasionaliste. Unpublished master's thesis. Stellenbosch: Stellenbosch University.

Van Niekerk, A.A. 1983. Die grense van die kritiese Rede. *Tydskrif vir Geesteswetenskappe*, 23(1):14-29.

Verkiel, S.E. 2017. Amoral enhancement. *Journal of Medical Ethics*. 43:52-55.

Walker, M. 2009. Enhancing genetic virtue: a project for twenty-first century humanity? *Politics and the Life Sciences*, 28(2):27-47.

Walker, M. 2010. In defense of the genetic virtue program. *Politics and the Life Science*s, 29(1):90-96.

Wallace, B., Cesarini, D., Lichtenstein, P. & Johannesson, M. 2007. Heritability of Ultimatum Game Responder Behaviour. *Proceedings of the National Academy of Sciences*, 104(40):15631-4.

Wasserman, D. 2014. When bad people do good things: will moral enhancement make the world a better place?. *Journal of Medical Ethics*, 40(6):374-375.

Watson, G. 1975. Free Agency. *Journal of Philosophy*, 72(8):205-220.

Wilson, A.T. 2014. Egalitarianism and successful moral bioenhancement. *American Journal of Bioethics*, 14(4):35-36.

Wiseman, H. 2014. Moral enhancement – "hard" and "soft" forms. *American Journal of Bioethics*, 14(4):48-49.

Wittgenstein, L. 2001. *Philosophical Investigations*. G.E.M. Anscombe (tr.). Oxford: Basil Blackwell.

Young, L. & Duncan, J. 2012. Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience*, 7(1):1-10.

Zarpentine, C. 2013. The thorny and arduous path of moral progress: moral psychology and moral enhancement. *Neuroethics*, 6(1):141-153.